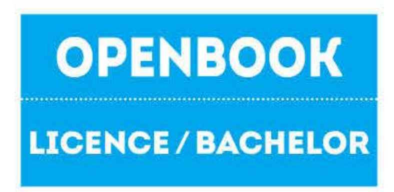


STATISTIQUE ET PROBABILITÉS EN ÉCONOMIE-GESTION

CHRISTOPHE HURLIN, VALÉRIE MIGNON

DUNOD

Les contenus complémentaires et les corrigés des exercices sont disponibles en ligne sur www.dunod.com à la page du livre.

Conseiller éditorial : Lionel Ragot

Création graphique de la maquette intérieure : SG Créations
Création graphique de la couverture : Hokus Pokus Créations

Illustrations : Judith Chouraqui

Crédits iconographiques : p. 84 : Chee-Onn Leong – Fotolia.com ; p. 254 : Kashisu – Fotolia.com ; p. 290 : lenets_tan – Fotolia.com ; couverture : Master Lu – Fotolia.com

5 rue Laromiguière, 75005 Paris
www.dunod.com

ISBN 978-2-10-072532-8

Sommaire

Avant-propos

Qu'est-ce que la statistique ? La statistique est une science recouvrant plusieurs dimensions. On emploie d'ailleurs très fréquemment le pluriel « statistiques » pour désigner cette discipline et témoigner ainsi de sa diversité. La statistique englobe la recherche et la collecte de données, leur traitement et leur analyse, leur interprétation, leur présentation sous la forme de tableaux et graphiques, le calcul d'indicateurs permettant de les caractériser et synthétiser... Ces différents éléments renvoient à ce que l'on a coutume de nommer la statistique descriptive, fondée sur l'observation de données relatives à toutes sortes de phénomènes (économiques, financiers, historiques, géographiques, biologiques, etc.).

Il arrive cependant fréquemment que les données représentatives du phénomène que l'on souhaite étudier ne soient pas parfaitement connues, c'est-à-dire pas toutes parfaitement observables, au sens où elles ne fournissent qu'une information partielle sur l'ensemble du phénomène que l'on analyse. Afin de pouvoir en réaliser une étude statistique, il est alors nécessaire d'inférer des informations à partir des quelques éléments dont on dispose. En d'autres termes, le statisticien devra effectuer des hypothèses concernant les lois de probabilité auxquelles obéit le phénomène à analyser. La statistique fait alors appel à la théorie des probabilités et est qualifiée de statistique mathématique ou encore de statistique inférentielle.

Un bref retour sur l'histoire. Même si le terme de « statistique » est généralement considéré comme datant du XVIIIe siècle[1], le recours à cette discipline remonte à un passé bien plus éloigné. On fait en effet souvent référence à la collecte de données en Chine en 2238 av. J.-C. concernant les productions agricoles, ou encore en Égypte en 1700 av. J.-C. en référence au cadastre et au cens. La collecte de données à des fins descriptives est ainsi bien ancienne, mais ce n'est qu'au XVIIIe siècle qu'est apparue l'idée d'utiliser les statistiques à des fins prévisionnelles. Ce fut le cas en démographie où les statistiques collectées lors des recensements de la population ont permis l'élaboration de tables de mortalité en Suède et en France.

Du côté des mathématiciens, les recherches sur le calcul des probabilités se sont développées dès le XVIIe siècle, au travers notamment des travaux de Fermat et Pascal. Même si Condorcet et Laplace ont proposé quelques exemples d'application de la théorie des probabilités, ce n'est qu'au cours de la deuxième moitié du XIXe siècle, grâce aux travaux de Quételet, que l'apport du calcul des probabilités à la statistique fut réellement mis en évidence, conduisant ainsi aux prémisses de la statistique mathématique. Cette dernière s'est ensuite largement développée à la fin du XIXe siècle et dans la première moitié du XXe siècle.

Par la suite, grâce notamment aux progrès de l'informatique peu avant la deuxième moitié du XXe siècle, de nouvelles méthodes d'analyse ont vu le jour, comme l'analyse multidimensionnelle permettant d'étudier de façon simultanée plusieurs types de données. La deuxième moitié du XXe siècle est aussi la période durant laquelle plusieurs courants de pensée en statistique s'affrontent, notamment autour de la notion de probabilité.

1 On attribue en effet ce terme au professeur allemand Gottfried Achenwall (1719-1772) qui, en 1746, emploie le mot *Statistik* dérivé de *Staatskunde*.

Les domaines d'application de la statistique sont multiples. Initialement employée en démographie, elle est en effet utilisée dans toutes les sciences humaines et sociales comme l'économie, la finance, la gestion, le marketing, l'assurance, l'histoire, la sociologie, la psychologie, etc., mais aussi en médecine, en sciences de la terre et du vivant (biologie, géologie...), météorologie, etc. Cet éventail des domaines illustre ainsi toute la richesse de la statistique dont cet ouvrage vise à rendre compte.

En quoi ce manuel se distingue-t-il des autres ouvrages de statistique ?

Tout en présentant de façon rigoureuse tous les développements théoriques nécessaires, cet ouvrage propose un exposé clair et pédagogique des différents concepts en les illustrant par de très nombreux exemples et cas concrets. Le lecteur sera ainsi à même de répondre à de multiples questions qui se posent au quotidien dans les domaines de l'économie, la finance et la gestion.

Chaque chapitre débute par des questions et exemples concrets, permettant de mettre en avant l'intérêt des concepts statistiques qui vont être étudiés. Afin de répondre à ces interrogations et traiter ces cas concrets, les différents outils et méthodes statistiques sont ensuite présentés. L'exposé est ainsi progressif, mêlant de façon harmonieuse définitions littéraire et mathématique. En fin de chapitre figurent des exercices qui permettent au lecteur d'évaluer et tester les connaissances acquises. Les exercices font l'objet de **corrigés très détaillés, disponibles en ligne sur www.dunod.com**, sur la page de l'ouvrage. Le lecteur trouvera également sur cette page Internet des annexes à télécharger reproduisant les principales **tables statistiques**, ainsi que de nombreux **compléments** relatifs à plusieurs chapitres de l'ouvrage.

Diverses rubriques spécifiques à la collection « Openbook » composent les chapitres. Outre les prérequis et les objectifs propres à chaque chapitre, une rubrique « Les grands auteurs » présente de façon synthétique un auteur clé dont les travaux ont profondément marqué le développement de la statistique. La rubrique « Focus » permet quant à elle de faire rapidement le point sur un concept fondamental, alors que la rubrique « Pour aller plus loin » offre la possibilité au lecteur d'approfondir un ou plusieurs points particuliers. La rubrique « En pratique » permet également au lecteur de se familiariser avec l'application concrète d'un concept ou d'une méthode. Enfin, la rubrique « Trois questions à... » illustre l'orientation résolument appliquée de l'ouvrage en donnant la parole à quelques grands acteurs du monde professionnel, nous expliquant la façon dont ils utilisent la statistique au quotidien.

Comment est organisé ce manuel ? Cet ouvrage a pour objectif de fournir au lecteur l'ensemble des connaissances que doit acquérir un étudiant au cours de son cursus de licence en économie-gestion ou de son cycle d'études Bac+3. Il couvre donc les trois années du cycle Bac+3 (licence ou bachelor). Il s'organise ainsi en trois parties, chacune étant relative à une année du cycle Bac+3. La première partie, correspondant au programme de la première année post-bac, traite de la statistique descriptive et comporte quatre chapitres. Le chapitre 1 étudie les distributions à un caractère et présente l'ensemble des concepts de base de la statistique descriptive : tableaux, graphiques et caractéristiques clés comme la moyenne, la variance, la médiane, etc. Le chapitre 2 étend l'analyse au cas de deux variables statistiques et porte ainsi sur les distributions à deux caractères. Le chapitre 3 offre une présentation des indices, très

utilisés en pratique. Le chapitre 4 propose quant à lui une introduction à l'analyse des séries temporelles en dotant le lecteur de l'ensemble des outils nécessaires pour l'analyse de l'évolution d'un phénomène au cours du temps.

La deuxième partie de l'ouvrage, correspondant au programme de la deuxième année du cycle Bac+3, relève du domaine de la statistique mathématique et se compose également de quatre chapitres. La notion fondamentale de probabilité fait l'objet du chapitre 5. Le chapitre 6 traite des variables aléatoires, c'est-à-dire des variables dont les valeurs sont soumises au hasard. L'étude de ces variables nécessite le recours à des lois de probabilité, dont les plus usuelles (lois normale, binomiale, de Student, de Poisson...) sont présentées au cours du chapitre 7. Le chapitre 8 clôt la deuxième partie par l'étude des propriétés de convergence.

La troisième partie de l'ouvrage, correspondant au programme de la dernière année du cycle Bac+3, traite de l'estimation et des tests. Le chapitre 9 est relatif à l'estimation, le chapitre 10 proposant quant à lui une description de l'une des méthodes les plus utilisées connue sous le nom de maximum de vraisemblance. La théorie des tests statistiques fait l'objet du chapitre 11, dernier chapitre du manuel.

Remerciements. Cet ouvrage est le fruit de divers enseignements de statistique dispensés par les auteurs en première, deuxième et troisième années de licence à l'Université d'Orléans et à l'Université Paris Ouest–Nanterre La Défense. Nous adressons nos remerciements à nos étudiants dont les questions et commentaires lors de nos cours ont naturellement contribué à la présentation pédagogique de ce manuel. Nous remercions Lionel Ragot pour la confiance qu'il nous a accordée en nous encourageant à rédiger ce manuel, ainsi que les éditions Dunod. Nous remercions très vivement nos collègues et amis Cécile Couharde, Olivier Darné, Emmanuel Dubois, Gilles Dufrénot, Elena Dumitrescu, Meglena Jeleva et Hélène Raymond pour leur relecture très attentive et pour leurs remarques et suggestions toujours très constructives. Emmanuel Dubois nous a également aidé pour la réalisation de certains graphiques dans la première partie de l'ouvrage, qu'il en soit chaleureusement remercié. Alina Catargiu, Axelle Chauvet-Peyrard, Andreea Danci, Damien Deballon, Laurent Ferrara, Yoann Grondin, Abdou Ndiaye, Ekaterina Sborets et Stéphanie Tring ont très gentiment accepté de répondre à nos questions, nous leur adressons nos plus vifs remerciements pour leurs contributions. Enfin, nous remercions très sincèrement nos familles pour leur soutien sans faille et leur patience lors de la rédaction de cet ouvrage.

À Séverine, Josiane, Emmanuel et Pierre.
À Tania et Emmanuel.

Table des matières

Partie 1

Statistique descriptive

Initialement employée en démographie dans le cadre des recensements de la population, la statistique descriptive est utilisée dans de nombreux domaines et disciplines, comme l'économie, la finance, l'assurance, le marketing, l'histoire, la géographie, la géologie, la biologie, la médecine, la météorologie, le sport, etc. Ce très large éventail de domaines d'application s'explique par le fait que dès lors que l'on dispose de données, c'est-à-dire d'observations, sur le phénomène que l'on souhaite étudier, il est nécessaire de les traiter afin de pouvoir les exploiter pour en extraire un certain nombre d'informations pertinentes. Tel est précisément l'objet de la statistique descriptive, qui permet de résumer et synthétiser l'ensemble des données étudiées au travers de graphiques, tableaux et divers indicateurs dont l'un des plus connus est la moyenne.

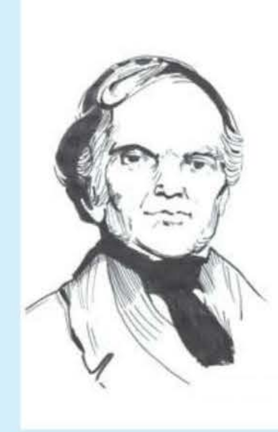
Au-delà de l'analyse d'un seul phénomène, la statistique descriptive permet aussi d'analyser et chiffrer la relation entre plusieurs phénomènes, c'est-à-dire plusieurs variables, et de mesurer l'intensité d'une telle liaison.

Chapitre 1

Quel est le salaire annuel moyen des hommes et des femmes en France ? Quelle est la proportion d'hommes et de femmes gagnant plus que ce salaire moyen ? À quel niveau de salaire se situe la plus grande partie de la population ? Les salaires ont-ils beaucoup fluctué ces cinquante dernières années ? Ont-ils suivi une évolution similaire pour les hommes et les femmes ? Les femmes sont-elles victimes d'inégalités salariales ?

La **statistique descriptive** permet de répondre à toutes ces questions. Elle permet en effet de résumer et synthétiser, par le biais de tableaux, graphiques et indicateurs statistiques, l'ensemble des données étudiées.

LES GRANDS AUTEURS



William Playfair (1759-1823)

Ingénieur et économiste écossais, **William Playfair** est considéré comme l'un des pionniers de la représentation graphique des données statistiques. Dans son ouvrage *Commercial and Political Atlas* paru en 1786, il introduit plusieurs représentations graphiques, comme celle retraçant l'évolution temporelle des intérêts de la dette publique britannique au cours du XVIII[e] siècle ou encore le **diagramme en bâtons** lui permettant de comparer les importations et exportations de l'Écosse en 1781 à celles d'autres pays. Également crédité de l'invention du célèbre **histogramme**, les représentations graphiques proposées par Playfair figurent parmi celles les plus utilisées en statistique descriptive. Quelques années plus tard, son ouvrage *Statistical Breviary* paru en 1801 présente un schéma circulaire, connu aujourd'hui sous le nom de **représentation par secteurs** (ou « camembert »). ■

Distributions à un caractère

Plan

Pré-requis

→ **Connaître** les opérations mathématiques de base.

Objectifs

→ **Synthétiser, résumer et extraire** l'information pertinente contenue dans une série statistique.
→ **Représenter** graphiquement une distribution statistique.
→ **Construire** un tableau statistique.
→ **Définir** les indicateurs statistiques clés.

Le tableau 1.1 donne la valeur du salaire annuel net moyen en euros des hommes et des femmes en France de 1950 à 2010 (source des données : INSEE). La figure 1.1 représente graphiquement ces mêmes données : la courbe orange décrit l'évolution du salaire des hommes sur la période 1950-2010, la courbe grise étant relative à l'évolution du salaire des femmes sur la même période. Sans prendre en compte l'effet de l'inflation, on constate globalement une tendance haussière avec un niveau plus élevé du salaire pour les hommes que pour les femmes.

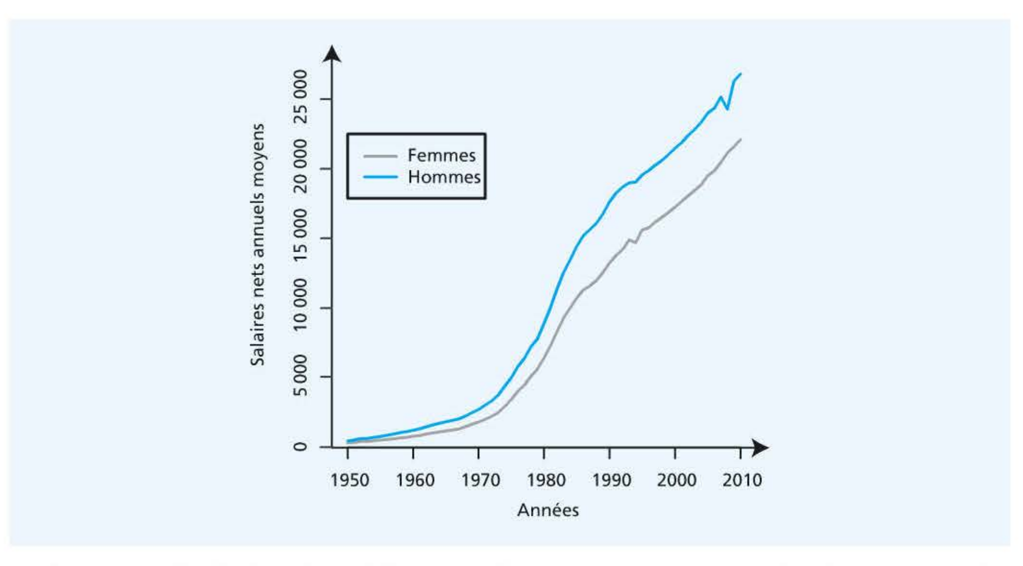
▼ **Tableau 1.1 Salaire annuel net moyen en euros en France, 1950-2010**

Année	Femmes	Hommes	Année	Femmes	Hommes	Année	Femmes	Hommes
1950	310	444	1970	1 807	2 711	1990	13 258	17 643
1951	344	530	1971	2 002	3 020	1991	13 772	18 266
1952	402	622	1972	2 218	3 330	1992	14 225	18 708
1953	412	637	1973	2 487	3 746	1993	14 894	18 999
1954	462	694	1974	2 946	4 388	1994	14 703	19 054
1955	504	771	1975	3 424	5 009	1995	15 606	19 580
1956	550	854	1976	4 009	5 799	1996	15 782	19 896
1957	600	947	1977	4 465	6 380	1997	16 187	20 278
1958	669	1 051	1978	5 102	7 223	1998	16 506	20 607
1959	711	1 122	1979	5 616	7 804	1999	16 861	21 033
1960	789	1 227	1980	6 418	8 881	2000	17 259	21 498
1961	849	1 327	1981	7 298	10 041	2001	17 651	21 889
1962	941	1 460	1982	8 343	11 411	2002	18 072	22 422
1963	1 037	1 604	1983	9 287	12 587	2003	18 443	22 840
1964	1 099	1 714	1984	9 996	13 464	2004	18 858	23 360
1965	1 168	1 820	1985	10 718	14 430	2005	19 500	24 007
1966	1 240	1 935	1986	11 302	15 212	2006	19 866	24 370
1967	1 316	2 036	1987	11 590	15 639	2007	20 472	25 168
1968	1 479	2 231	1988	11 991	16 093	2008	21 135	24 287
1969	1 648	2 473	1989	12 561	16 776	2009	21 593	26 300
						2010	22 112	26 831

Source : INSEE.

De tels tableaux et graphiques visent ainsi à résumer et rendre lisible l'information contenue dans les données étudiées (ici le salaire). Ils doivent être complétés par le calcul de divers indicateurs statistiques qui nous permettront notamment de déterminer le niveau moyen du salaire sur la période considérée, le niveau du salaire tel que le nombre d'individus (hommes et femmes) percevant moins que ce niveau est identique au nombre d'individus gagnant plus, le niveau du salaire perçu par le plus grand nombre des individus étudiés, ou encore la dispersion, c'est-à-dire la variabilité, du salaire entre hommes et femmes et/ou au cours de la période d'étude. À cette fin, on

calcule des indicateurs dits de tendance centrale, de forme et de dispersion. Le recours aux indicateurs de concentration nous permet en outre de compléter l'analyse afin de quantifier précisément les inégalités de salaires entre hommes et femmes.



▲ **Figure 1.1** **Évolution du salaire annuel net moyen en euros des hommes et des femmes en France, de 1950 à 2010**

1 Définitions et concepts fondamentaux de la statistique descriptive

L'objectif de la statistique descriptive est de résumer et synthétiser l'information contenue dans les données étudiées afin d'en déduire un certain nombre de propriétés. À cette fin, on utilise des tableaux et des graphiques (▸ section 1.2) et l'on calcule divers indicateurs ou caractéristiques (▸ section 2).

1.1 Définitions

1.1.1 Population, individus, échantillon

Une **population** est un ensemble, fini ou non, d'éléments que l'on souhaite étudier. Ces éléments portent le nom d'**individus** ou d'**unités statistiques**. Il peut s'agir par exemple d'êtres humains (adultes, enfants, chômeurs, salariés, etc.), d'animaux ou encore d'objets (entreprises, voitures, ordinateurs, incendies, accidents, etc.). Très souvent, la population que l'on souhaite analyser est très grande et il est usuel de se restreindre à l'étude d'un échantillon.

Un **échantillon** est ainsi un sous-ensemble de la population considérée qui doit posséder les mêmes caractéristiques statistiques que la population dont il est issu. À partir d'un échantillon dit **représentatif**, il est alors possible d'effectuer des analyses et d'en déduire des conclusions valables pour la population.

1.1.2 Caractères, modalités et variables statistiques

Caractères et modalités. Afin d'étudier les individus composant une population, on les classe en un certain nombre de sous-ensembles, appelés **caractères** ou **variables statistiques**. À titre d'exemple, si l'on étudie le personnel salarié d'une entreprise, on pourra retenir comme caractères le sexe, l'âge, la profession, le salaire, l'ancienneté dans l'entreprise, etc. Pour une voiture, on retiendra la puissance du moteur, le nombre de places assises, la couleur, le modèle... Les valeurs possibles prises par le caractère ou la variable sont appelées **modalités**. La variable « sexe » a ainsi deux modalités, masculin et féminin, mais les caractères peuvent avoir un très grand nombre de modalités. Notons que les modalités doivent être incompatibles – un individu ne peut pas appartenir simultanément à plusieurs modalités – et exhaustives – toutes les situations possibles doivent être recensées.

Une variable peut être **qualitative** ou **quantitative**. Dans le premier cas, les modalités ne sont pas des valeurs chiffrées, elles ne sont pas mesurables mais uniquement observables (sexe, nationalité, catégorie socio-professionnelle, etc.). Dans le cas d'une variable quantitative, les modalités sont mesurables : à chaque modalité est associé un nombre, c'est-à-dire une valeur chiffrée, représentant la mesure du caractère. Ainsi, la puissance d'un moteur, le nombre de places assises, l'âge, la taille, etc. sont des variables statistiques dont les modalités sont des nombres.

Variables statistiques qualitatives nominales et ordinales. Les variables qualitatives peuvent être **nominales** ou **ordinales**. Dans le premier cas, les modalités ne peuvent être ordonnées, contrairement au cas de variables ordinales. Des exemples usuels de variables nominales sont le sexe (modalités : masculin, féminin), l'état civil (modalités : célibataire, marié ou pacsé, veuf, divorcé), la couleur des yeux ou encore le groupe sanguin. Des variables comme le niveau d'études (avec, par exemple, comme modalités : sans diplôme, primaire, secondaire, universitaire) ou le niveau de satisfaction (peu satisfait, satisfait, très satisfait) sont des variables ordinales. Notons toutefois que le fait de pouvoir ordonner ou non les modalités d'une variable peut être sujet à débats. Prenons l'exemple de la variable « catégorie socio-professionnelle ». Si l'on a coutume d'ordonner comme suit les trois modalités « ouvriers », « employés », « cadres », il devient plus difficile d'ordonner les modalités « enseignant », « chercheur » et « responsable administratif » (en particulier si ces trois modalités correspondent au même niveau de diplôme et/ou de responsabilités).

Variables statistiques quantitatives discrètes et continues et regroupement en classes. Les variables quantitatives peuvent être discrètes ou continues. Une variable est dite **discrète** lorsque ses valeurs sont des nombres isolés dans son intervalle de variation. Il s'agit en règle générale de nombres entiers ; par exemple le nombre d'enfants par famille, le nombre de salariés d'une entreprise, le nombre d'automobiles vendues. Une variable est dite **continue** lorsqu'elle peut prendre toutes

les valeurs au sein de son intervalle de variation. On peut donner comme exemples la taille, le poids, la température, etc. Le nombre de valeurs possibles à l'intérieur de l'intervalle de variation étant infini, on les groupe par **classes**. Si l'on considère la variable de salaire annuel, on peut par exemple définir les classes suivantes : moins de 10 000 euros, de 10 000 à moins de 15 000 euros, de 15 000 à moins de 20 000 euros, de 20 000 à moins de 25 000 euros, de 25 000 à moins de 40 000 euros, plus de 40 000 euros. La longueur (ou l'étendue) de la classe, c'est-à-dire la différence entre l'extrémité supérieure et l'extrémité inférieure de la classe, est appelée **amplitude de la classe**. Elle peut être variable, comme dans l'exemple précédent, ou constante. Dans la mesure où il existe une infinité de valeurs au sein d'une classe, il est possible de calculer le **centre de classe** défini comme suit :

$$\text{Centre de classe} = \frac{\text{Extrémité inférieure} + \text{Extrémité supérieure}}{2} \quad (1.1)$$

EN PRATIQUE

La distinction variables discrètes/variables continues

Du fait de la précision limitée des mesures, il peut être difficile de distinguer entre variables discrètes et continues. On retient en conséquence fréquemment le groupement ou non en classes comme moyen de distinction : une variable continue est ainsi souvent telle que le nombre de ses valeurs est si important qu'il convient de les regrouper en classes afin de pouvoir l'étudier.

S'agissant des classes, mentionnons (i) que le nombre d'individus par classe doit être suffisamment important de sorte à limiter ou éliminer les variations accidentelles qui peuvent se produire si l'on retient un effectif trop faible et (ii) que les amplitudes ne doivent pas être trop importantes afin de conserver certaines particularités de la variable étudiée.

1.1.3 Fréquences et effectifs

Considérons une population comprenant N individus. Ce nombre est appelé **effectif total** de la population. On regroupe les N individus suivant les k modalités, notées $x_i, i = 1,\ldots,k$, de la variable x. À chaque modalité correspond un nombre d'individus $n_i, i = 1,\ldots,k$, appelé **effectif** (ou fréquence absolue)[1] de la modalité x_i. Dans le cas d'une variable quantitative ou qualitative ordinale, la somme des effectifs n_i pour $i = 1,\ldots,k$ est ainsi égale à l'effectif total de la population :

$$N = \sum_{i=1}^{k} n_i \quad (1.2)$$

La **fréquence** (ou fréquence relative) associée à une modalité x_i est définie comme le rapport :

$$f_i = \frac{n_i}{N} \quad (1.3)$$

1 Dans le cas d'une variable qualitative nominale, l'effectif n_i correspond au nombre de fois où la modalité x_i apparaît.

La fréquence donne la proportion d'individus de la population présentant la modalité x_i et est en général exprimée en pourcentage. En utilisant l'équation (1.2), on déduit immédiatement la propriété suivante :

$$\sum_{i=1}^{k} f_i = 1 = 100\ \% \tag{1.4}$$

La somme des fréquences f_i correspondant aux différentes modalités, notée F_i, est appelée **fréquence cumulée** :

$$F_1 = f_1 \tag{1.5}$$

$$F_2 = f_1 + f_2 \tag{1.6}$$

$$\ldots$$

$$F_i = f_1 + f_2 + \ldots + f_j + \ldots + f_i \tag{1.7}$$

soit :

$$F_i = \sum_{j=1}^{i} f_j \tag{1.8}$$

La fréquence cumulée F_i indique la proportion des individus pour lesquels la variable étudiée est strictement inférieure à x_{i+1}.

On définit de la même façon les effectifs cumulés :

$$N_i = \sum_{j=1}^{i} n_j \tag{1.9}$$

1.2 Tableaux statistiques et représentations graphiques

Les individus classés suivant les caractères et modalités forment une **distribution** (ou une **série**) statistique qui peut être synthétisée sous la forme de tableaux statistiques et de graphiques : une série représente ainsi la suite des valeurs prises par la variable étudiée. Ces tableaux sont à une dimension si l'on ne considère qu'un seul caractère et à deux dimensions si l'on retient deux caractères (► chapitre 2).

FOCUS

Variable statistique et variable aléatoire

Ainsi que nous l'avons vu, une variable est une entité pouvant prendre toutes les valeurs possibles au sein d'un ensemble de définition donné. Lorsque les valeurs prises par la variable sont soumises au hasard (par exemple, « pile » ou « face » dans le cas du lancer d'une pièce), on parle de **variable aléatoire** (► chapitre 6). Il convient de ne pas les confondre avec les **variables statistiques**, objet d'étude de ce premier chapitre. La distribution d'une variable statistique est une distribution *empirique*. Les différentes caractéristiques qui seront présentées dans ce chapitre se réfèrent à cette distribution empirique : fonction de répartition *empirique*, moyenne *empirique*, variance *empirique*, moments *empiriques*, etc. Dans la suite du chapitre, afin d'alléger la présentation nous omettrons généralement le terme « empirique », mais il convient de bien garder cette notion à l'esprit.

1.2.1 Distributions à caractère qualitatif

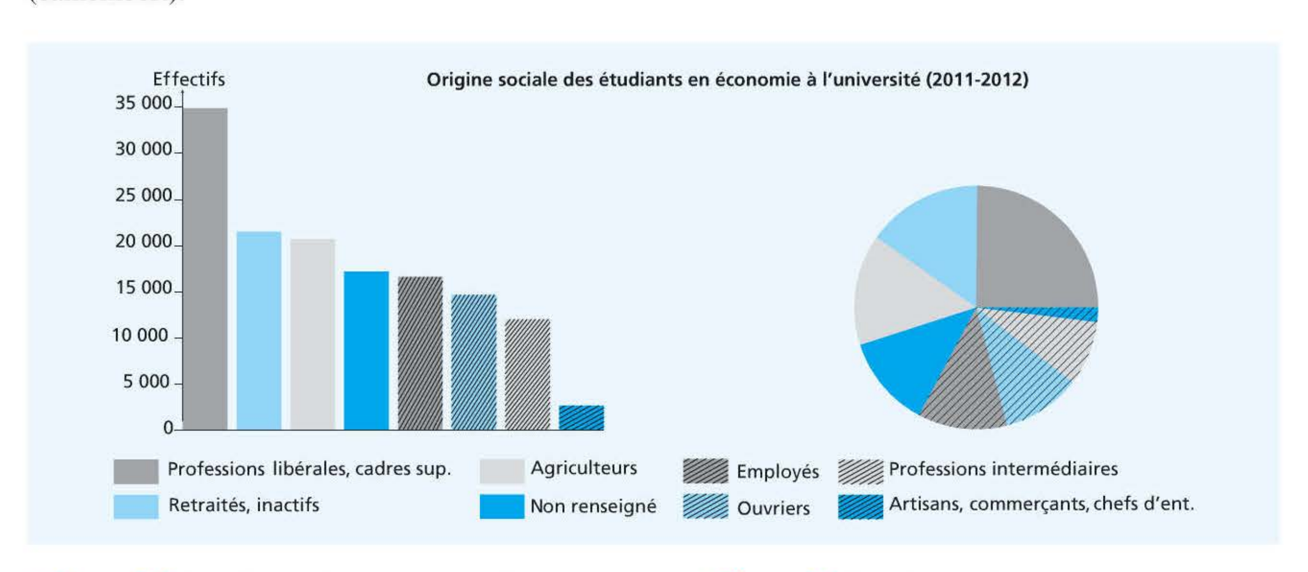
Considérons l'origine sociale des étudiants en économie durant l'année universitaire 2011-2012. Le tableau 1.2 reporte, dans la première colonne, les 8 modalités considérées. Les deuxième et troisième colonnes donnent respectivement l'effectif pour chaque modalité et la fréquence correspondante ; cette dernière étant égale au rapport entre l'effectif de chaque modalité et l'effectif total (140 205 étudiants). On constate ainsi que près de 25 % des étudiants en économie ont leurs parents cadres supérieurs ou exerçant une profession libérale. Une très faible proportion, 1,9 %, d'étudiants est issue du milieu agricole.

▼ **Tableau 1.2** Origine sociale des étudiants en économie à l'université en 2011-2012

Modalités	Effectifs	Fréquences
Agriculteurs	2 665	1,9
Artisans, commerçants, chefs d'entreprise	12 029	8,6
Professions libérales, cadres supérieurs	34 867	24,9
Professions intermédiaires	14 666	10,5
Employés	17 186	12,3
Ouvriers	16 601	11,8
Retraités, inactifs	21 506	15,3
Non renseigné	20 685	14,8
Total	**140 205**	**100,0**

Source : Ministère de l'Enseignement Supérieur et de la Recherche, MESR (DGESIP-DGRI-SIES).

Deux principaux types de graphiques sont utilisés pour des distributions à caractère qualitatif : la **représentation en tuyaux d'orgue** et la **représentation par secteurs** (camembert).



▲ **Figure 1.2** Représentation en tuyaux d'orgue ▲ **Figure 1.3** Représentation par secteurs

Dans les deux cas, le principe de base est que les surfaces doivent être proportionnelles aux effectifs. Sur le graphique 1.2 en tuyaux d'orgue, les différentes modalités sont représentées par des rectangles de base constante et de hauteurs proportionnelles aux effectifs. Il est également possible de considérer les fréquences au lieu des effectifs en ordonnée. Dans le cas d'une représentation par secteurs (▶ figure 1.3), l'effectif total est représenté par un cercle et les modalités par des secteurs dont la surface (et donc l'angle au centre) est proportionnelle à l'effectif.

1.2.2 Distributions à caractère quantitatif

Cas des variables discrètes. Considérons la répartition du nombre d'enfants sur un échantillon de 150 familles. La première colonne du tableau 1.3 reporte les différentes modalités (nombre d'enfants par famille), la deuxième colonne les effectifs pour chacune des modalités, la troisième colonne la fréquence correspondante, la dernière colonne donnant la fréquence cumulée. On constate ainsi que 31,33 % des familles ont moins de 2 enfants, 61,33 % des familles ont moins de 3 enfants, et ainsi de suite.

De façon générale, le tableau statistique d'une variable discrète sera de la forme représentée dans le tableau 1.4.

▼ Tableau 1.3 Nombre d'enfants par famille

Modalités	Effectifs	Fréquences	Fréquences cumulées
0	10	6,67	6,67
1	37	24,67	31,33
2	45	30	61,33
3	24	16	77,33
4	16	10,67	88,00
5	9	6	94,00
6	6	4	98,00
7	3	2	100,00
Total	**150**	**100**	

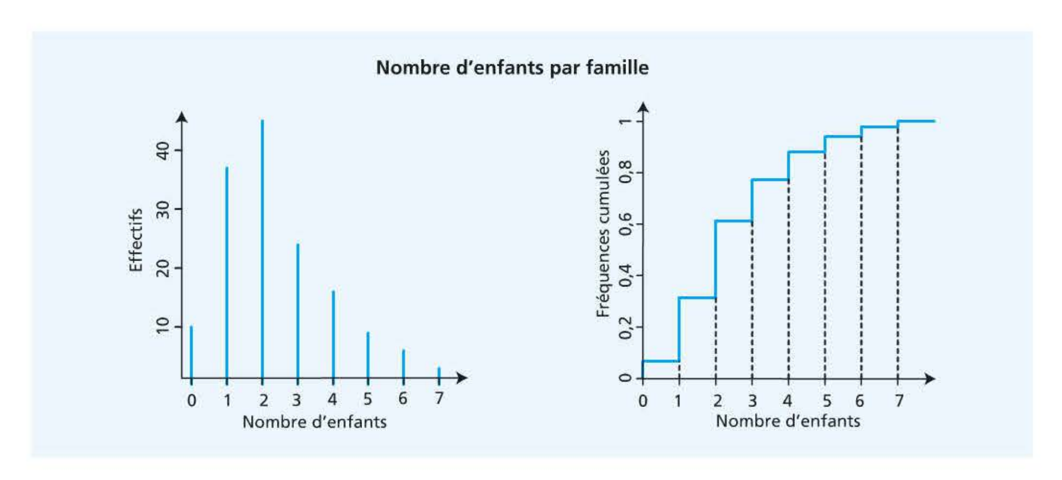
Deux types de graphiques existent pour les variables quantitatives discrètes : le **diagramme en bâtons** et le **diagramme cumulatif** (ou diagramme intégral). Dans un diagramme en bâtons, on fait correspondre à chaque valeur des modalités x_i (en abscisse) un bâton vertical de longueur proportionnelle à l'effectif n_i ou à la fréquence f_i associée (en ordonnée). La figure 1.4 reporte ainsi le diagramme en bâtons correspondant aux données du tableau 1.3. Notons que dans le cas où ce sont les fréquences qui sont reportées en ordonnée, la courbe joignant les sommets des bâtons est appelée **courbe des fréquences**.

Le diagramme cumulatif (ou courbe cumulative) consiste à représenter les fréquences cumulées (ou, de façon similaire, les effectifs cumulés) sur un graphique en escalier (▶ figure 1.5) [2]. Les valeurs des modalités x_i de la variable x étudiée figurent en abs-

2 La courbe joignant les extrémités droites des « marches d'escalier » est appelée **courbe des fréquences cumulées**.

▼ **Tableau 1.4** Tableau statistique d'une variable quantitative discrète

Modalités x_i	Effectifs n_i	Fréquences $f_i = n_i/N$	Fréquences cumulées $F_i = \sum_{j=1}^{i} f_j$
x_1	n_1	f_1	F_1
x_2	n_2	f_2	$F_2 = f_1 + f_2$
⋮	⋮	⋮	⋮
x_i	n_i	f_i	$F_i = f_1 + f_2 + ... + f_i$
⋮	⋮	⋮	⋮
x_k	n_k	f_k	$F_k = f_1 + f_2 + ... + f_k = 1$
Total	N	**1 (ou 100 %)**	



▲ **Figure 1.4** Diagramme en bâtons ▲ **Figure 1.5** Courbe cumulative

cisse, la hauteur de chaque marche de l'escalier étant proportionnelle à la fréquence cumulée correspondante. Le diagramme cumulatif représente ainsi la proportion, notée $F_x(x_i)$, des individus de l'échantillon pour lesquels la valeur de la variable x est inférieure à x_i. Cette fonction, définie pour toute valeur de x, est appelée fonction cumulative ou **fonction de répartition** (empirique)[3] et est donnée par :

$$F_x(x_i) = \sum_{j=1}^{i} f_j \tag{1.10}$$

Si l'on reprend le tableau 1.3, il est ainsi aisé de constater que plus de 60 % (61,33 %) des familles ont moins de 3 enfants.

Cette fonction est telle que :

$$\lim_{x_i \to +\infty} F_x(x_i) = 1 \quad \text{et} \quad \lim_{x_i \to -\infty} F_x(x_i) = 0 \tag{1.11}$$

3 Rappelons qu'il s'agit d'une fonction de répartition *empirique* puisqu'elle se rapporte à une variable statistique (et non pas à une variable aléatoire comme ce sera le cas dans le chapitre 6).

Cas des variables continues. Considérons la répartition des enfants scolarisés par âge, de 2 ans à moins de 22 ans, durant l'année 2010-2011 en France. S'agissant d'une variable continue, les données sont regroupées en classes et sont reportées dans le tableau 1.5.

▼ **Tableau 1.5 Répartition des enfants scolarisés par âge en 2010-2011 en France**

Numéro de classe i	Classes	Effectifs n_i	Fréquences f_i	Fréquences cumulées F_i
1	2 à moins de 6 ans	2 538 643	18,36	18,36
2	6 à moins de 10 ans	3 220 753	23,29	41,65
3	10 à moins de 14 ans	3 174 548	22,96	64,61
4	14 à moins de 18 ans	2 967 358	21,46	86,07
5	18 à moins de 22 ans	1 925 926	13,93	100
Total		**13 827 228**	**100**	

Source : Ministère de l'Éducation nationale (MEN), MESR, INSEE.

De façon générale, en notant e_{i-1} la borne (ou extrémité) inférieure de la classe i et e_i la borne supérieure de cette même classe, le tableau statistique d'une variable continue prend la forme de celui représenté dans le tableau 1.6.

▼ **Tableau 1.6 Tableau statistique d'une variable quantitative continue**

Numéro de classe i	Classes $[e_{i-1}, e_i[$	Effectifs n_i	Fréquences $f_i = n_i/N$	Fréquences cumulées $F_i = \sum_{j=1}^{i} f_j$
1	$[e_0, e_1[$	n_1	f_1	F_1
2	$[e_1, e_2[$	n_2	f_2	$F_2 = f_1 + f_2$
⋮	⋮	⋮	⋮	⋮
i	$[e_{i-1}, e_i[$	n_i	f_i	$F_i = f_1 + f_2 + ... + f_i$
⋮	⋮	⋮	⋮	⋮
k	$[e_{k-1}, e_k[$	n_k	f_k	$F_k = f_1 + f_2 + ... + f_k = 1$
Total		N	**1 (ou 100 %)**	

Dans la mesure où une variable quantitative continue peut prendre une infinité de valeurs au sein d'une classe donnée, la représentation graphique en diagramme en bâtons n'est pas appropriée. Pour représenter une variable quantitative continue, on utilise un **histogramme** : à chaque classe de la variable, portée en abscisse, on associe un rectangle ayant pour base l'amplitude de la classe et dont la hauteur est proportionnelle à l'effectif (ou à la fréquence). On doit distinguer le cas où les classes ont toutes la même amplitude du cas d'amplitudes différentes. Considérons tout d'abord le cas, comme celui décrit dans le tableau 1.5, où les classes ont toutes la même amplitude, soit ici 4 ans. Comme illustré par l'histogramme reporté sur la figure 1.6, la hauteur de chaque rectangle est proportionnelle à la fréquence f_i. On obtient naturellement un graphique similaire si l'on remplace les fréquences f_i par les effectifs n_i.



▲ **Figure 1.6 Répartition des enfants scolarisés par âge en 2010-2011 en France, histogramme**

Remarque : La courbe joignant le milieu des sommets des rectangles est appelée courbe ou **polygone des fréquences**. Une telle courbe est notamment utilisée lorsque l'échantillon comprend un très grand nombre d'individus, rendant la représentation en histogramme peu lisible du fait des regroupements des observations en un nombre relativement faible de classes.

Considérons à présent le cas où les classes n'ont pas la même amplitude. Reprenons et complétons à cette fin l'exemple de la répartition des enfants scolarisés en France en considérant une classe supplémentaire, la classe allant de 22 ans à moins de 30 ans (▸ tableau 1.7).

▼ **Tableau 1.7 Répartition des enfants scolarisés par âge en 2010-2011 en France**

Numéro de classe i	Classes	Effectifs n_i	Fréquences f_i	Amplitude a_i	Amplitude a'_i	Hauteur h_i
1	[2,6[	2 538 643	17,31	4	1	17,31
2	[6,10[	3 220 753	21,96	4	1	21,96
3	[10,14[	3 174 548	21,64	4	1	21,64
4	[14,18[	2 967 358	20,23	4	1	20,23
5	[18,22[	1 925 926	13,13	4	1	13,13
6	[22,30[	840 518	5,73	8	2	2,87
Total		**14 667 746**	**100**			

Source : MEN, MESR, INSEE.

Ainsi que nous le constatons dans le tableau 1.7, l'amplitude a_i des 5 premières classes est de 4 ans, la dernière classe ayant quant à elle une amplitude de 8 ans. Pour pouvoir comparer les effectifs ou les fréquences des différentes classes, il convient de « corriger » les amplitudes afin que l'aire de chaque rectangle composant l'histogramme soit bien proportionnelle à l'effectif (ou la fréquence). À cette fin, on choisit une amplitude

unité a_u, qui est en général l'amplitude la plus fréquente ou la plus faible. Ici, nous retenons donc une amplitude unité égale à 4 ans. On exprime les amplitudes de chaque classe en fonction de cette nouvelle unité. Soient a'_i les amplitudes ainsi corrigées :

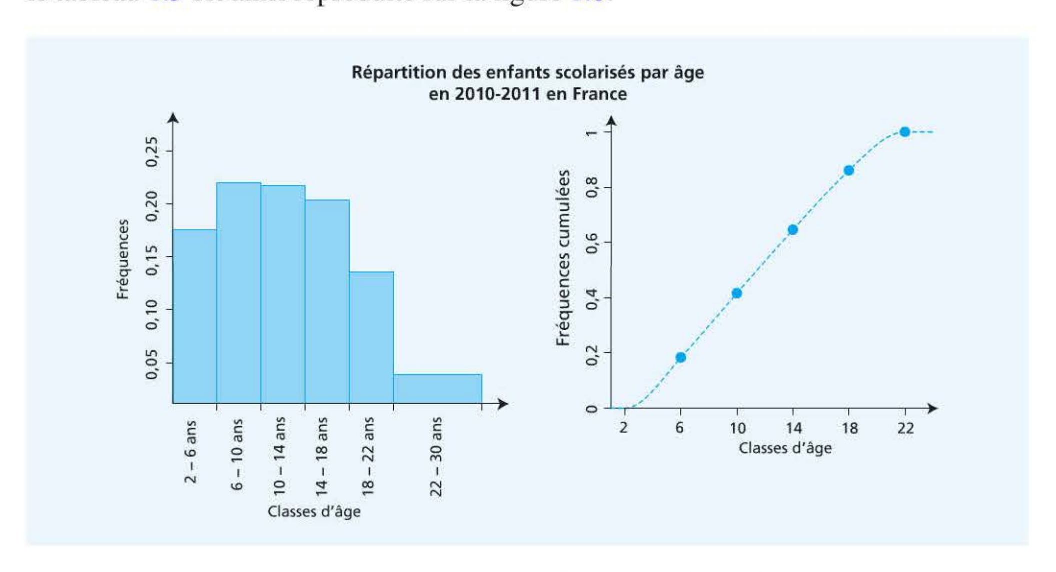
$$a'_i = \frac{e_i - e_{i-1}}{a_u} \tag{1.12}$$

Il suffit ensuite de calculer la hauteur h_i des rectangles comme suit :

$$h_i = \frac{f_i}{a'_i} \tag{1.13}$$

et l'on peut alors tracer l'histogramme dans lequel l'aire de chaque rectangle est bien proportionnelle à la fréquence (ou l'effectif) de la classe correspondante (▸ figure 1.7).

L'obtention de la fonction de répartition empirique d'une variable continue est similaire au cas d'une variable discrète et cette fonction vérifie les mêmes propriétés aux limites. La fonction de répartition empirique correspondant aux données figurant dans le tableau 1.5 est ainsi reproduite sur la figure 1.8.



▲ Figure 1.7 Histogramme

▲ Figure 1.8 Courbe cumulative

2 Caractéristiques d'une distribution à un caractère

Ainsi que nous l'avons vu dans la section précédente, les tableaux et graphiques nous permettent de disposer d'une première description des données étudiées. Un graphique nous donne une idée de l'ordre de grandeur de la variable considérée, au travers des valeurs de la variable situées au centre de la distribution. On parle alors de **tendance centrale**. Un graphique nous fournit également une indication quant à la variabilité des données autour de cette tendance centrale, on parle alors de **dispersion**. Pour mesurer la tendance centrale et la dispersion, il convient de calculer des caractéristiques permettant de décrire plus précisément la distribution que les graphiques. On y adjoint des caractéristiques de **forme** et de **concentration**.

FOCUS

Les conditions de Yule

Les caractéristiques doivent remplir un certain nombre de propriétés, appelées **conditions de Yule**. Une caractéristique doit ainsi :

– être objective, c'est-à-dire indépendante de l'observateur ;

– utiliser l'information de façon exhaustive, c'est-à-dire être basée sur l'ensemble des observations de la série ;

– être facilement interprétable et calculable ;

– être peu sensible aux fluctuations d'échantillonnage ;

– se prêter aisément au calcul algébrique.

2.1 Caractéristiques de tendance centrale

2.1.1 Mode

Définition 1.1

Le **mode** d'une distribution est la valeur de la variable qui correspond à l'effectif ou à la fréquence le (la) plus élevé(e). Il s'agit donc de la valeur la plus fréquemment rencontrée dans une distribution.

Le mode peut être calculé pour tous les types de variables (qualitative et quantitative).

Cas d'une variable discrète. Reprenons le tableau 1.3 ou, de façon équivalente, la figure 1.4. Le mode est la modalité pour laquelle la fréquence est la plus élevée, c'est-à-dire pour laquelle la bâton est le plus haut sur le graphique. Il s'agit donc ici de la valeur 2, ce qui signifie que la majorité des familles considérées ont 2 enfants.

Notons que lorsque la série étudiée comporte deux valeurs consécutives pour lesquelles la fréquence est la plus élevée, on parle d'*intervalle modal* – les bornes de cet intervalle correspondant à ces deux valeurs de la série. Mentionnons en outre que lorsque la distribution étudiée ne comporte qu'un seul mode – ce qui est le cas le plus fréquent – on parle de *distribution unimodale*. Il peut toutefois arriver que la distribution comporte 2 ou plusieurs modes (correspondant à 2 ou plusieurs valeurs non consécutives), on parle alors de *distributions bi-modale* ou *pluri-modale*. La présence de plusieurs modes est indicative d'une certaine hétérogénéité de l'échantillon analysé.

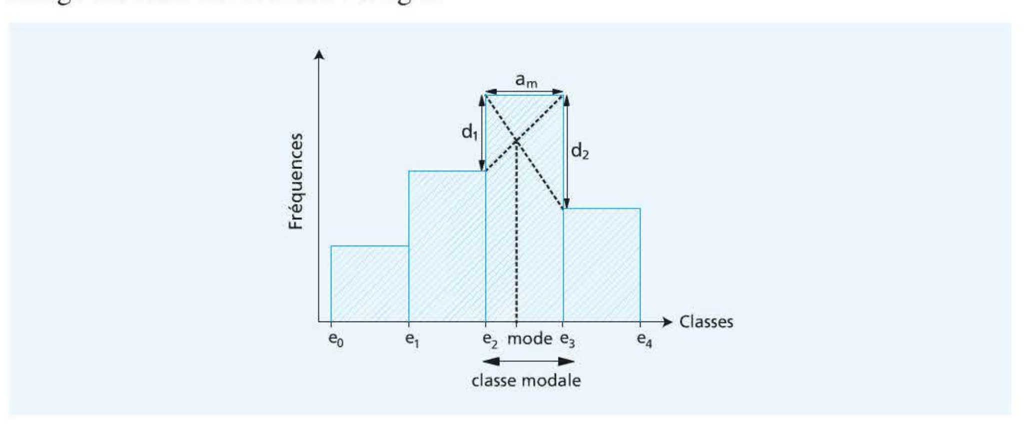
Cas d'une variable continue. Les données étant regroupées en classes, on détermine la **classe modale** qui correspond à la classe du tableau ou de l'histogramme pour laquelle la fréquence est la plus élevée. Dans le cas de l'exemple relatif à la répartition par âge des enfants scolarisés (▶ tableau 1.5), la classe modale est la classe [6,10[. Ainsi que l'illustre la figure 1.9, il est possible de déterminer la valeur précise du mode :

$$Mode = e_{i-1} + a_m \times \frac{d_1}{d_1 + d_2} \qquad (1.14)$$

où e_{i-1} désigne la valeur de l'extrémité inférieure de la classe modale, a_m l'amplitude de cette même classe, d_1 la différence entre l'effectif de la classe modale et l'effectif de la classe précédente et d_2 la différence entre l'effectif de la classe modale et l'effectif de la classe suivante. Dans le cadre de notre exemple, la valeur du mode est donnée par :

$$Mode = 6 + 4 \times \frac{(3\,220\,753 - 2\,538\,643)}{(3\,220\,753 - 2\,538\,643) + (3\,220\,753 - 3\,174\,548)} = 9{,}75 \quad (1.15)$$

Remarque : Dans le cas où les classes sont d'amplitudes différentes, il convient de corriger les effectifs ou les fréquences préalablement à la détermination du mode en utilisant la procédure présentée dans la section 1.2.2. Dans la formule (1.14), d_1 et d_2 désignent alors des effectifs corrigés.



▲ **Figure 1.9** Détermination du mode, cas d'une variable continue

2.1.2 Médiane

Définition 1.2

La **médiane** est la valeur de la variable qui partage la série étudiée en deux sous-ensembles d'effectifs égaux.

En d'autres termes, la médiane – qui peut être calculée sur des variables quantitatives ou qualitatives ordinales – est telle que le nombre des individus ayant une valeur inférieure soit égal au nombre des individus ayant une valeur supérieure. Il s'agit de la valeur M de la variable pour laquelle la fréquence cumulée est égale à 1/2 :

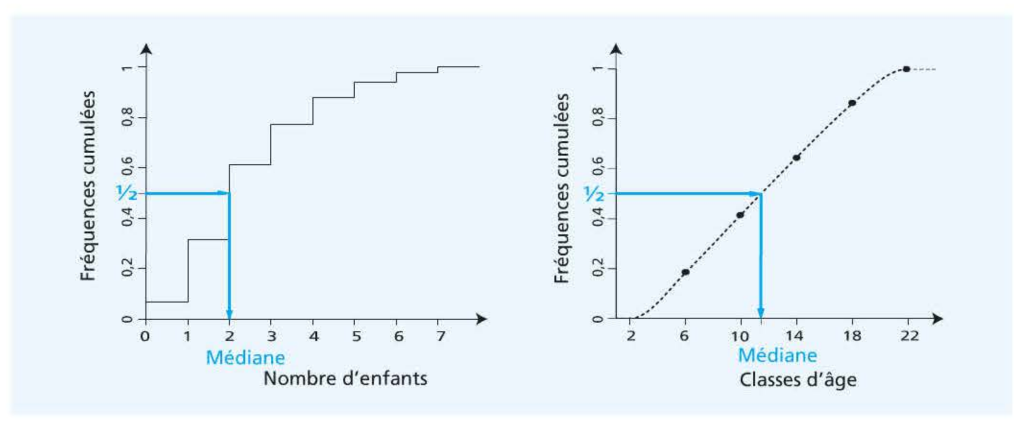
$$F_x(M) = \frac{1}{2} \quad (1.16)$$

Exemple

Si la note médiane des étudiants à l'examen de mathématiques en première année est égale à 12/20, cela signifie qu'il y a autant d'étudiants ayant obtenu moins de 12/20 que d'étudiants ayant obtenu plus de 12/20.

Cas d'une variable discrète. La détermination de la médiane nécessite au préalable de classer les observations de la série par ordre croissant. Considérons par exemple 9 étudiants ayant obtenu les notes suivantes (sur 20) : 4 ; 3 ; 17 ; 19 ; 11 ;

10 ; 12 ; 14 ; 13. On classe ces observations par ordre croissant, soit : 3 ; 4 ; 10 ; 11 ; 12 ; 13 ; 14 ; 17 ; 19. La médiane est égale à 12/20 : il y a autant d'étudiants ayant obtenu une note en dessous de 12 que d'étudiants ayant obtenu plus de 12/20. On constate que la médiane est très simple à calculer dans le cadre de cet exemple qui comprend un nombre impair d'observations. Supposons désormais que l'on ajoute à notre échantillon un dixième étudiant dont la note s'élève à 8/20. La série ordonnée s'écrit donc : 3 ; 4 ; 8 ; 10 ; 11 ; 12 ; 13 ; 14 ; 17 ; 19. Dans ce cas, on définit un intervalle médian, donné par [11, 12], la médiane étant quant à elle égale à (11 + 12)/2 = 11,5. Plus généralement, reprenons le cas du nombre d'enfants par famille (▶ tableau 1.3). Ainsi que nous l'avons vu, la médiane est la valeur pour laquelle la fréquence cumulée est égale à 1/2. Cette valeur n'apparaît pas dans le tableau 1.3, elle se situe entre les modalités « 1 enfant » ($F_1 = 0{,}3133$) et « 2 enfants » ($F_2 = 0{,}6133$). Par convention, on retient dans ce cas comme valeur médiane la valeur correspondant à la ligne la plus basse des deux dans le tableau, soit 2 enfants par famille. Cette détermination de la médiane est représentée graphiquement sur la figure 1.10.



▲ **Figure 1.10 Nombre d'enfants par famille, détermination de la médiane**

▲ **Figure 1.11 Répartition des enfants scolarisés par âge en 2010-2011 en France, détermination de la médiane**

Cas d'une variable continue. Contrairement au cas discret, la médiane peut toujours être exactement déterminée dans le cas d'une variable continue et est obtenue à partir des fréquences cumulées. Reprenons les données du tableau 1.5 relatives à la scolarisation par âge. La proportion d'enfants scolarisés ayant moins de 10 ans est de 41,65 %, celle d'enfants scolarisés ayant moins de 14 ans est égale à 64,61 %. La valeur de la médiane est donc comprise entre 10 et 14 ans. La classe [10,14[est appelée **classe médiane**. La valeur de la médiane peut être déterminée graphiquement grâce à la fonction de répartition, ainsi que cela est reproduit sur la figure 1.11.

Numériquement, cette valeur peut s'obtenir aisément par interpolation linéaire, *via* la relation suivante :

$$M = e_{i-1} + \frac{a_i}{f_i} \times [0{,}5 - F_{i-1}] \qquad (1.17)$$

où e_{i-1} est l'extrémité inférieure de la classe médiane (10 dans notre exemple), a_i l'amplitude de la classe médiane (4), f_i la fréquence de la classe médiane (22,96 %) et F_{i-1} désigne la fréquence cumulée de la classe au dessus de la classe médiane dans

le tableau (41,65 %). On en déduit :

$$M = 10 + \frac{4}{0{,}2296} \times [0{,}5 - 0{,}4165] \simeq 11{,}45 \quad (1.18)$$

Il s'ensuit que notre échantillon est composé d'un nombre identique d'enfants scolarisés ayant moins de 11,45 ans que d'enfants scolarisés ayant plus de 11,45 ans.

Remarque : La présence de classes d'amplitudes inégales n'affecte pas le calcul de la médiane. Il n'est donc pas nécessaire de corriger les effectifs ou les fréquences pour déterminer la médiane.

2.1.3 Quantiles

Les quantiles sont des valeurs permettant de partager les observations ordonnées d'une série en sous-groupes contenant le même nombre de données (aux erreurs d'arrondis près).

Définition 1.3

Le **quantile** d'ordre q est défini par :

$$F_x(x_q) = q \quad (1.19)$$

avec $0 \leq q \leq 1$.
Lorsque $q = 1/2$, on retrouve la médiane, cette dernière étant un quantile particulier. On distingue trois principaux types de quantiles :

– Les **quartiles** : ce sont les valeurs de la variable qui partagent la distribution en 4 sous-ensembles égaux. Il existe donc 3 quartiles : $Q_1 = 0{,}25$; $Q_2 = 0{,}5$ et $Q_3 = 0{,}75$; le deuxième quartile Q_2 étant la médiane. L'intervalle $Q_3 - Q_1$ est appelé intervalle interquartile. Il comprend 50 % des observations et est en général utilisé comme caractéristique de dispersion (voir *infra*).
– Les **déciles** : ce sont les valeurs de la variable qui partagent la distribution en 10 sous-ensembles égaux. Il existe 9 déciles, notés $D_1, \ldots, D_9$. L'intervalle $D_9 - D_1$, appelé intervalle interdécile, comprend 80 % des observations et est également utilisé comme caractéristique de dispersion (voir *infra*).
– Les **centiles** : ce sont les valeurs de la variable qui partagent la distribution en 100 sous-ensembles égaux. En notant C_1 et C_{99} les premier et dernier centiles, respectivement, on définit l'intervalle intercentile $C_{99} - C_1$ comprenant 98 % des observations.

2.1.4 Moyenne arithmétique

Définition 1.4

La **moyenne arithmétique**[4] d'une variable quantitative x, notée $\bar{x}$, est égale à la somme des valeurs, $x_1, x_2, \ldots, x_N$, prises par cette variable divisée par le nombre

4 Rappelons qu'il s'agit de la moyenne arithmétique *empirique* au sens où elle se rapporte à une variable statistique (et non pas à une variable aléatoire comme dans le chapitre 6).

d'observations N, soit :

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (1.20)$$

À titre d'exemple, considérons un échantillon de 12 étudiants ayant obtenu les notes suivantes (sur 20) : 4 ; 3 ; 17 ; 19 ; 11 ; 10 ; 12 ; 14 ; 13 ; 12 ; 4 ; 10. La note moyenne est donc égale à 10,75/20 :

$$\bar{x} = \frac{4 + 3 + 17 + 19 + 11 + 10 + 12 + 14 + 13 + 12 + 4 + 10}{12} = \frac{129}{12} = 10{,}75 \quad (1.21)$$

Il est également possible de regrouper au sein d'un tableau les observations ayant une valeur identique, en indiquant l'effectif correspondant (► tableau 1.8).

▼ **Tableau 1.8** **Moyenne arithmétique pondérée**

Note x_i	Effectifs n_i	$n_i x_i$
3	1	3
4	2	8
10	2	20
11	1	11
12	2	24
13	1	13
14	1	14
17	1	17
19	1	19
Total	**12**	**129**

Pour calculer la moyenne, il convient alors de pondérer les notes par les effectifs correspondants, soit :

$$\bar{x} = \frac{3 \times 1 + 4 \times 2 + 10 \times 2 + 11 \times 1 + 12 \times 2 + 13 \times 1 + 14 \times 1 + 17 \times 1 + 19 \times 1}{12} \quad (1.22)$$

d'où :

$$\bar{x} = \frac{129}{12} = 10{,}75 \quad (1.23)$$

Il s'agit d'une moyenne arithmétique pondérée dont la définition est donnée ci-après.

Définition 1.5

La **moyenne arithmétique pondérée** d'une variable x composée des observations $x_1, x_2, \ldots, x_k$ auxquelles sont associés les effectifs $n_1, n_2, \ldots, n_k$ est donnée par :

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_k x_k}{N} = \frac{1}{N}\sum_{i=1}^{k} n_i x_i \quad (1.24)$$

On peut également exprimer la moyenne arithmétique pondérée en fonction des fréquences f_i, $i = 1,\ldots,k$:

$$\bar{x} = \frac{n_1}{N}x_1 + \frac{n_2}{N}x_2 + \ldots + \frac{n_k}{N}x_k \qquad (1.25)$$

$$\bar{x} = f_1x_1 + f_2x_2 + \ldots + f_kx_k \qquad (1.26)$$

$$\bar{x} = \sum_{i=1}^{k} f_ix_i \qquad (1.27)$$

Dans le cas d'une variable discrète, il est possible d'appliquer directement la formule donnée par l'équation (1.24). On ajoute ainsi une colonne donnant le produit n_ix_i comme dans le tableau 1.8 et l'on obtient aisément :

$$\bar{x} = \frac{129}{12} = 10{,}75 \qquad (1.28)$$

Dans le cas d'une variable continue, les observations étant groupées en classes, il convient de déterminer préalablement le centre de classe x_i et d'appliquer ensuite la formule donnée par l'équation (1.24). À titre d'exemple, le tableau 1.9 reporte le salaire mensuel de 150 salariés d'une entreprise de textile, « TextpluS ».

▼ **Tableau 1.9** **Salaire mensuel de 150 salariés, en euros**

Classes de salaires	Effectifs n_i	Centres de classes x_i	n_ix_i
[1 400,1 600[	26	1 500	39 000
[1 600,1 800[	34	1 700	57 800
[1 800,2 000[	65	1 900	123 500
[2 000,2 200[	8	2 100	16 800
[2 200,2 400[	10	2 300	23 000
[2 400,2 600[	7	2 500	17 500
Total	**150**		**277 600**

Le calcul du salaire moyen donne en conséquence :

$$\bar{x} = \frac{277\,600}{150} = 1\,850{,}67 \qquad (1.29)$$

Le salaire mensuel moyen des salariés de l'entreprise considérée est ainsi égal à 1 850,67 euros.

Propriétés

La moyenne arithmétique vérifie deux propriétés importantes :

– La somme des écarts des observations à la moyenne $(x_i - \bar{x})$ est nulle :

$$\sum_{i=1}^{k} n_i(x_i - \bar{x}) = 0 \qquad (1.30)$$

– La somme des carrés des écarts des observations à la moyenne est inférieure à la somme des carrés des écarts par rapport à toute autre valeur.

Démonstration

Démontrons tout d'abord la relation (1.30). On a :

$$\sum_{i=1}^{k} n_i(x_i - \bar{x}) = \sum_{i=1}^{k} n_i x_i - \sum_{i=1}^{k} n_i \bar{x} = \sum_{i=1}^{k} n_i x_i - \bar{x} \sum_{i=1}^{k} n_i \qquad (1.31)$$

Or $\sum_{i=1}^{k} n_i = N$ et $\sum_{i=1}^{k} n_i x_i = N\bar{x}$. On en déduit donc la relation (1.30) :

$$\sum_{i=1}^{k} n_i(x_i - \bar{x}) = N\bar{x} - \bar{x}N = 0 \qquad (1.32)$$

Pour montrer à présent que la somme des carrés des écarts des observations à la moyenne est inférieure à la somme des carrés des écarts par rapport à toute autre valeur, considérons une valeur quelconque X et minimisons la somme des carrés des écarts par rapport à cette valeur, notée $S(X)$:

$$S(X) = \sum_{i=1}^{k} n_i(x_i - X)^2 \qquad (1.33)$$

En annulant la dérivée première de $S(X)$ par rapport à X, il vient :

$$\sum_{i=1}^{k} n_i(x_i - X) = 0 \qquad (1.34)$$

soit :

$$\sum_{i=1}^{k} n_i x_i - \sum_{i=1}^{k} n_i X = 0 \qquad (1.35)$$

ce que l'on peut encore écrire, en notant que $\sum_{i=1}^{k} n_i X = X \sum_{i=1}^{k} n_i$ et $\sum_{i=1}^{k} n_i = N$:

$$\sum_{i=1}^{k} n_i x_i - NX = 0 \qquad (1.36)$$

D'où :

$$X = \frac{1}{N} \sum_{i=1}^{k} n_i x_i = \bar{x} \qquad (1.37)$$

La somme des carrés des écarts $S(X)$ est donc bien minimale pour $X = \bar{x}$. ■

POUR ALLER PLUS LOIN

► Voir p. 31

2.2 Caractéristiques de dispersion

Étudier la dispersion – ou la variabilité – d'une série consiste à analyser ses fluctuations autour d'une valeur centrale.

2.2.1 Étendue

Définition 1.6

L'**étendue** est définie comme la différence entre la plus grande et la plus petite valeur de la série.

L'étendue est ainsi très simple à calculer, mais présente l'inconvénient majeur de ne dépendre que des valeurs extrêmes de la série. Ces dernières étant souvent exceptionnelles, voire aberrantes, l'étendue ne constitue qu'une mesure très imparfaite de la dispersion d'une série.

2.2.2 Intervalles interquantiles

Contrairement à l'étendue, les **intervalles interquantiles** permettent d'exclure les valeurs extrêmes de la série en ne retenant qu'un certain pourcentage du nombre total d'observations. L'intervalle le plus fréquemment utilisé est l'intervalle interquartile $Q_3 - Q_1$ contenant 50 % des observations de la série. Malgré son avantage de simplicité de calcul évident, son principal inconvénient réside dans le fait qu'il ne tient compte que de l'ordre des observations et non pas de leurs valeurs. Il s'agit en conséquence, tout comme l'étendue, d'une mesure imparfaite de la dispersion.

2.2.3 Écart absolu moyen

L'**écart absolu moyen**, noté EAM, permet de mesurer la dispersion d'une série *via* la moyenne des valeurs absolues des écarts de chaque observation par rapport à la moyenne :

$$EAM = \frac{1}{N}\sum_{i=1}^{k} n_i|x_i - \bar{x}| \tag{1.38}$$

2.2.4 Variance et écart-type

La variance et l'écart-type sont les mesures de dispersion les plus utilisées[5].

Définition 1.7

La **variance**, notée $V(x)$, d'une variable statistique x est donnée par la moyenne arithmétique des carrés des écarts des observations x_i, $i = 1,\ldots,k$, à la moyenne :

$$V(x) = \frac{1}{N}\sum_{i=1}^{k} n_i(x_i - \bar{x})^2 \tag{1.39}$$

L'**écart-type**, noté σ_x, est la racine carrée de la variance et s'exprime ainsi dans la même unité que la variable étudiée :

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{k} n_i(x_i - \bar{x})^2} \tag{1.40}$$

Plus l'écart-type est faible, plus les valeurs sont regroupées autour de la moyenne. Ainsi, si l'écart-type d'une série de notes des étudiants de première année est faible, cela signifie que la promotion est relativement homogène. Inversement, un écart-type élevé témoigne d'une forte dispersion au sein de la promotion.

5 Rappelons qu'il s'agit de la variance et de l'écart-type *empiriques* au sens où ils se rapportent à une variable statistique (et non pas à une variable aléatoire comme dans le chapitre 6).

L'utilisation des formules (1.39) et (1.40) s'avère relativement fastidieuse en pratique puisqu'elle nécessite le calcul des écarts $(x_i - \bar{x})$. Afin de rendre l'application plus aisée, on utilise la formule développée de la variance. Développons ainsi l'équation (1.39) :

$$V(x) = \frac{1}{N}\sum_{i=1}^{k} n_i[x_i^2 - 2x_i\bar{x} + \bar{x}^2] \tag{1.41}$$

soit :

$$V(x) = \frac{1}{N}\sum_{i=1}^{k} n_i x_i^2 - 2\bar{x}\frac{1}{N}\sum_{i=1}^{k} n_i x_i + \bar{x}^2\frac{1}{N}\sum_{i=1}^{k} n_i \tag{1.42}$$

Sachant que $\sum_{i=1}^{k} n_i = N$ et que $\frac{1}{N}\sum_{i=1}^{k} n_i x_i = \bar{x}$, on en déduit :

$$V(x) = \frac{1}{N}\sum_{i=1}^{k} n_i x_i^2 - \bar{x}^2 \tag{1.43}$$

L'équation (1.43) est appelée **formule développée de la variance**. Afin d'illustrer son application, reprenons l'exemple du salaire mensuel des 150 employés de l'entreprise TextpluS et complétons le tableau initial (► tableau 1.9) par l'ajout de deux colonnes x_i^2 et $n_i x_i^2$ (► tableau 1.10).

▼ **Tableau 1.10** **Salaire mensuel de 150 salariés, en euros**

Classes de salaires	Effectifs n_i	Centres de classes x_i	$n_i x_i$	x_i^2	$n_i x_i^2$
[1400,1600[	26	1 500	39 000	2 250 000	58 500 000
[1600,1800[	34	1 700	57 800	2 890 000	98 260 000
[1800,2000[	65	1 900	123 500	3 610 000	234 650 000
[2000,2200[	8	2 100	16 800	4 410 000	35 280 000
[2200,2400[	10	2 300	23 000	5 290 000	52 900 000
[2400,2600[	7	2 500	17 500	6 250 000	43 750 000
Total	**150**		**277 600**		**523 340 000**

L'application de la formule (1.43) donne :

$$V(x) = \frac{1}{150} \times 523\,340\,000 - \left(\frac{1}{150} \times 277\,600\right)^2 = 63\,966{,}22 \tag{1.44}$$

D'où :

$$\sigma_x = \sqrt{63\,966{,}22} = 252{,}92 \tag{1.45}$$

L'écart-type de la distribution des salaires est ainsi égal à 252,92 €. En pratique, lorsque l'on travaille sur un échantillon – et non sur une population – il convient de corriger les valeurs de la variance et de l'écart-type de ce que l'on nomme le biais de petit échantillon (► chapitre 9). On utilise les formules dites corrigées :

$$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{k} n_i(x_i - \bar{x})^2 = \frac{N}{N-1}V(x) \tag{1.46}$$

et :

$$S_x = \sqrt{S_x^2} = \sqrt{\frac{N}{N-1}}\sigma_x \quad (1.47)$$

où S_x^2 désigne la **variance corrigée** et S_x l'**écart-type corrigé**.

Propriétés

– La variance est toujours positive ou nulle : $V(x) \geq 0$.

– Le théorème de König-Huygens permet de relier la moyenne et la variance au travers de l'identité remarquable suivante :

$$\frac{1}{N}\sum_{i=1}^{k} n_i(x_i - \bar{x})^2 = \frac{1}{N}\sum_{i=1}^{k} n_i x_i^2 - \bar{x}^2 = \bar{x^2} - \bar{x}^2 \quad (1.48)$$

2.2.5 Coefficient de variation

Lorsque l'on souhaite comparer la dispersion de séries dont les unités sont différentes, par exemple la dispersion de salaires en euros et la dispersion de salaires en livres, il convient d'utiliser une mesure de dispersion relative. Le **coefficient de variation**, noté CV, fournit une telle mesure. Il s'agit d'un nombre sans dimension, indépendant des unités considérées, défini comme le rapport entre l'écart-type et la moyenne :

$$CV = \frac{\sigma_x}{\bar{x}} \quad (1.49)$$

Ainsi, si la valeur obtenue de CV sur une série de salaires en livres perçus par les salariés d'une entreprise A au Royaume-Uni est proche de celle obtenue sur une série de salaires en euros perçus par les salariés d'une entreprise B en France, on peut en déduire que la dispersion des salaires est proche dans les deux entreprises.

2.3 Caractéristiques de forme

2.3.1 Moments d'une distribution

Définition 1.8

On appelle **moment d'ordre** r ($r = 0,1,\ldots,N$) par rapport à une valeur quelconque a, la quantité notée M_r telle que[6] :

$$M_r = \frac{1}{N}\sum_{i=1}^{k} n_i(x_i - a)^r \quad (1.50)$$

où a est l'origine du moment.

On distingue les moments simples (ou ordinaires) et les moments centrés.

6 Rappelons qu'il s'agit dans ce chapitre des moments *empiriques* au sens où ils se rapportent à une variable statistique (et non pas à une variable aléatoire comme dans le chapitre 6).

Définition 1.9

On appelle **moment simple d'ordre** r un moment d'ordre r pour lequel $a = 0$, soit :

$$m_r = \frac{1}{N} \sum_{i=1}^{k} n_i x_i^r \quad (1.51)$$

On constate aisément que lorsque $r = 1$, le moment simple d'ordre 1 est la moyenne arithmétique.

Définition 1.10

On appelle **moment centré d'ordre** r un moment d'ordre r pour lequel $a = \bar{x}$, soit :

$$\mu_r = \frac{1}{N} \sum_{i=1}^{k} n_i (x_i - \bar{x})^r \quad (1.52)$$

Propriétés

– Le moment centré d'ordre 1 est nul :

$$\mu_1 = 0 \quad (1.53)$$

– Le moment centré d'ordre 2 est égal à la variance :

$$\mu_2 = V(x) \quad (1.54)$$

Démonstration

– Dans le cas où $r = 1$, on a :

$$\mu_1 = \frac{1}{N} \sum_{i=1}^{k} n_i (x_i - \bar{x}) \quad (1.55)$$

soit encore :

$$\mu_1 = \frac{1}{N} \sum_{i=1}^{k} n_i x_i - \frac{1}{N} \sum_{i=1}^{k} n_i \bar{x} = \bar{x} - \bar{x} \quad (1.56)$$

D'où :

$$\mu_1 = 0 \quad (1.57)$$

– Dans le cas où $r = 2$, on a :

$$\mu_2 = \frac{1}{N} \sum_{i=1}^{k} n_i (x_i - \bar{x})^2 \quad (1.58)$$

soit encore :

$$\mu_2 = \frac{1}{N} \sum_{i=1}^{k} n_i x_i^2 - \bar{x}^2 \quad (1.59)$$

D'où :

$$\mu_2 = m_2 - m_1^2 = V(x) \quad (1.60)$$

■

Remarque : Outre les moments centrés d'ordres 1 et 2, les moments centrés d'ordres 3 et 4 sont également fréquemment utilisés. Comme nous le verrons ci-après ils interviennent dans le calcul des coefficients d'asymétrie et d'aplatissement. Les expressions de ces deux moments sont les suivantes :

– Dans le cas où $r = 3$, on a : $\mu_3 = \frac{1}{N}\sum_{i=1}^{k} n_i(x_i - \bar{x})^3$

En développant cette égalité, on obtient l'expression du moment centré d'ordre 3 :

$$\mu_3 = m_3 + 2m_1^3 - 3m_1m_2 \qquad (1.61)$$

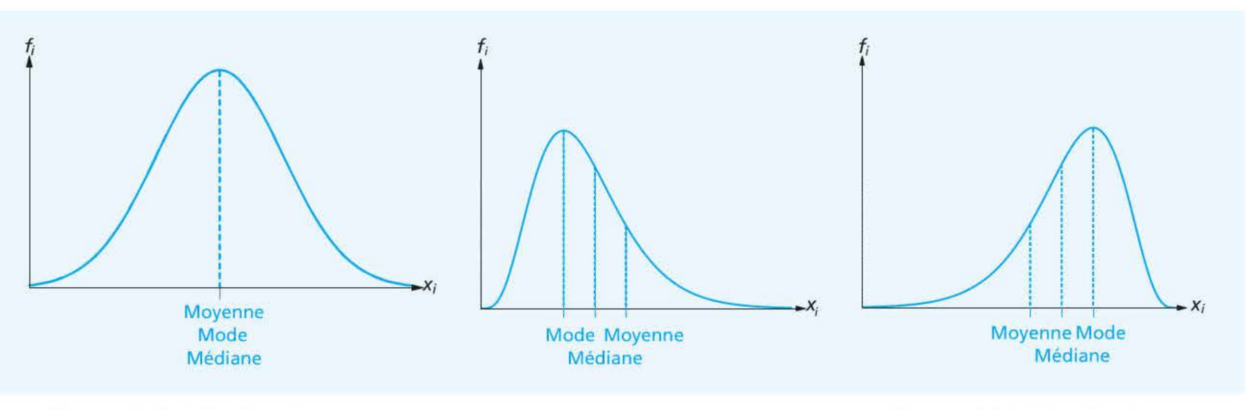
– Dans le cas où $r = 4$, on a : $\mu_4 = \frac{1}{N}\sum_{i=1}^{k} n_i(x_i - \bar{x})^4$

En développant cette égalité, on obtient l'expression du moment centré d'ordre 4 :

$$\mu_4 = m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4 \qquad (1.62)$$

2.3.2 Asymétrie et aplatissement

Asymétrie. L'**asymétrie** permet d'apprécier la répartition, régulière ou non, des observations autour d'une caractéristique de tendance centrale. Ainsi, lorsque les trois caractéristiques de tendance centrale (moyenne, mode et médiane) sont égales, la distribution (empirique) est dite symétrique (▶ figure 1.12). Lorsque qu'une distribution est telle que le mode est inférieur à la médiane, ces deux caractéristiques étant elles mêmes inférieures à la moyenne, la distribution est dite asymétrique et étalée vers la droite (ou oblique à gauche), comme cela est représenté sur la figure 1.13. Enfin, une distribution asymétrique étalée vers la gauche (ou oblique à droite) est telle que la moyenne est inférieure à la médiane, ces deux caractéristiques étant elles mêmes inférieures au mode (▶ figure 1.14).



▲ **Figure 1.12** **Distribution symétrique**

▲ **Figure 1.13** **Distribution étalée vers la droite**

▲ **Figure 1.14** **Distribution étalée vers la gauche**

Il est possible de quantifier l'asymétrie d'une distribution en calculant des coefficients. On distingue trois principaux **coefficients d'asymétrie** (ou coefficients de skewness) :

- Le **coefficient de Yule** : il consiste à comparer l'étalement à gauche et à droite de la distribution à l'aide des quartiles :

$$Yule = \frac{(Q_3 - M) - (M - Q_1)}{(Q_3 - M) + (M - Q_1)} \qquad (1.63)$$

où M est la médiane. Un coefficient de Yule nul correspond au cas de quartiles équidistants, signifiant que la distribution est symétrique. Lorsque que le coefficient de Yule est positif, la distribution est étalée vers la droite ; lorsqu'il est négatif, elle est étalée vers la gauche.

- Le **coefficient de Pearson**, noté β_1 : il est donné par[7] :

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \tag{1.64}$$

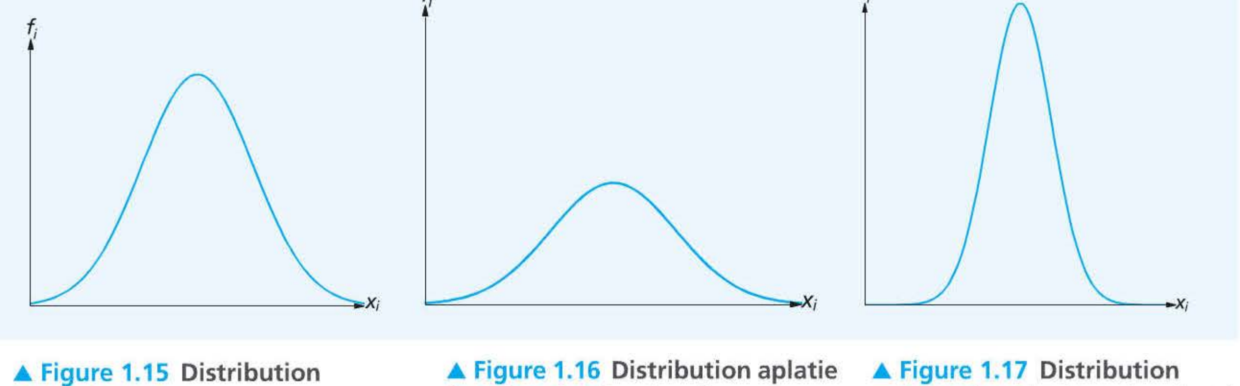
La distribution est symétrique si $\beta_1 = 0$, étalée vers la droite si $\beta_1 > 0$ et étalée vers la gauche si $\beta_1 < 0$.

- Le **coefficient de Fisher**, noté γ_1 : il est égal à la racine carrée de β_1 :

$$\gamma_1 = \sqrt{\beta_1} = \frac{\mu_3}{\sigma_x^3} \tag{1.65}$$

La distribution est symétrique si $\gamma_1 = 0$, étalée vers la droite si $\gamma_1 > 0$ et étalée vers la gauche si $\gamma_1 < 0$.

Aplatissement. L'**aplatissement** nous renseigne sur la relation entre la variation de la variable et la variation des fréquences. L'aplatissement s'étudie à partir de la courbe des fréquences que l'on compare à la distribution de la loi normale (▶ figure 1.15). Une courbe platykurtique correspond à une distribution aplatie (▶ figure 1.16), au sens où une forte variation de la variable entraine une faible variation de la fréquence et réciproquement. Une courbe leptokurtique (▶ figure 1.17) renvoie au cas d'une distribution pointue.



▲ **Figure 1.15** Distribution normale

▲ **Figure 1.16** Distribution aplatie (courbe platykurtique)

▲ **Figure 1.17** Distribution pointue (courbe leptokurtique)

7 Il existe également un autre coefficient d'asymétrie de Pearson, valable pour des distributions faiblement asymétriques, donné par : $\beta_1' = (\bar{x} - Mode)/\sigma_x$. La distribution est symétrique si $\beta_1' = 0$, étalée vers la droite si $\beta_1' > 0$ et étalée vers la gauche si $\beta_1' < 0$.

FOCUS

La loi normale

La **loi normale**, encore appelée loi de Gauss ou loi de Laplace-Gauss, est la loi statistique la plus répandue. Il s'agit d'une loi de probabilité continue très utilisée pour modéliser de nombreux phénomènes aléatoires. La courbe des fréquences de la loi normale est souvent dénommée « courbe en cloche » du fait de sa forme (▸ figure 1.15). Une distribution normale est caractérisée par un coefficient d'asymétrie nul, cette loi étant symétrique par rapport à la moyenne. Les coefficients d'aplatissement de Pearson et de Fisher d'une distribution normale sont respectivement égaux à 3 et 0.

Deux principaux **coefficients d'aplatissement** (ou coefficients de kurtosis) peuvent être calculés :

- Le coefficient de Pearson, noté β_2 : il est donné par :

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{V(x)^2} \tag{1.66}$$

  Pour une distribution normale, on a $\beta_2 = 3$. $\beta_2 < 3$ est caractéristique d'une courbe platykurtique et $\beta_2 > 3$ d'une courbe leptokurtique.

- Le coefficient de Fisher, noté γ_2 :

$$\gamma_2 = \beta_2 - 3 \tag{1.67}$$

  Pour une distribution normale, on a $\gamma_2 = 0$. $\gamma_2 < 0$ est caractéristique d'une courbe platykurtique et $\gamma_2 > 0$ d'une courbe leptokurtique.

2.4 Caractéristiques de concentration

Les caractéristiques de concentration sont plus particulièrement utilisées pour certaines distributions, comme l'étude des revenus, des salaires, des logements suivant leur surface, etc. Ces caractéristiques, qui s'appliquent au cas de variables continues à valeurs positives, permettent de mesurer des inégalités. Si l'on s'intéresse aux revenus, on considère qu'une société est parfaitement égalitaire si tous les individus ont le même salaire. À l'opposé, une société totalement inégalitaire est telle qu'un individu perçoit la totalité des revenus, les autres individus ne percevant aucun revenu.

2.4.1 Médiale

Définition 1.11

La **médiale** est la valeur x_i de x partageant la série du produit $n_i x_i$ en deux sous-ensembles égaux.

En d'autres termes, il s'agit de la médiane que l'on calcule non plus à partir des seuls effectifs n_i, mais à partir du produit $n_i x_i$. Si l'on suppose, à titre d'exemple, que la variable étudiée x est le salaire, $n_i x_i$ représente la masse salariale. En pratique, on utilise

la différence entre la médiale et la médiane comme un indicateur de concentration. Plus spécifiquement, soit ΔM cette différence :

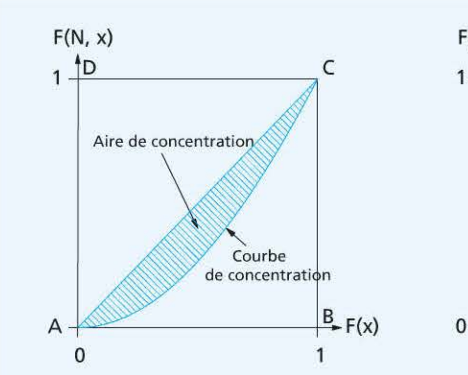
$$\Delta M = \text{Médiale} - \text{Médiane} \qquad (1.68)$$

On obtient $\Delta M = 0$ si la médiane est égale à la médiale, c'est-à-dire lorsque le système est parfaitement égalitaire au sens où tous les salariés perçoivent le même salaire (concentration nulle). Afin d'apprécier la concentration, on compare ΔM à la valeur de l'étendue. Si la valeur obtenue pour ΔM est supérieure à l'étendue, cela témoigne d'une forte concentration : une faible proportion de salariés perçoit une forte proportion de la masse salariale. À l'inverse, lorsque ΔM est inférieur à l'étendue, la concentration est faible.

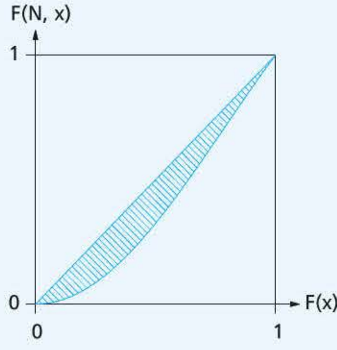
2.4.2 Courbe de concentration et indice de Gini

La **courbe de concentration** consiste à mettre en relation graphiquement les fréquences cumulées $F(x)$ de la série x_i (en abscisse) et les fréquences cumulées $F(N,x)$ de la série $n_i x_i$ (en ordonnée). La courbe s'inscrit dans un carré $ABCD$ représenté sur la figure 1.18 appelé **carré de Gini**. La première bissectrice correspond à une concentration nulle, c'est-à-dire à une parfaite équi-répartition.

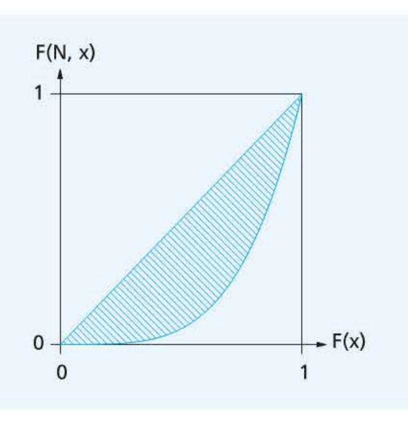
La courbe de concentration est également appelée **courbe de Lorenz** et l'aire située entre cette courbe et la première bissectrice correspond à l'**aire de concentration**. Ainsi, suivant la valeur de cette aire, on disposera d'un indicateur de concentration ; l'aire étant nulle si la médiane est égale à la médiale, c'est-à-dire si la concentration est nulle. Les figures 1.19 et 1.20 représentent schématiquement respectivement les cas de concentration faible et forte.



▲ **Figure 1.18** Courbe de concentration



▲ **Figure 1.19** Concentration faible



▲ **Figure 1.20** Concentration forte

Exemple

Reprenons l'exemple figurant dans le tableau 1.9 relatif au salaire mensuel (en euros) des 150 salariés de l'entreprise TextpluS.

La classe médiane est la classe [1 800,2 000[et l'on peut calculer la valeur précise de la médiane au moyen de la relation (1.17) :

$$M = 1800 + \frac{200}{0{,}43} \times [0{,}5 - 0{,}4] = 1\,846{,}51 \qquad (1.69)$$

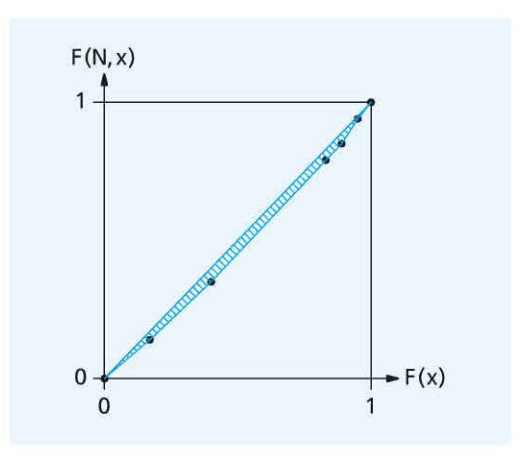
Il existe ainsi autant de salariés percevant moins de 1 846,51 euros que de salariés percevant plus que cette même somme. Au regard des valeurs figurant dans la colonne $\sum \frac{n_i x_i}{\sum n_i x_i}$ du tableau 1.11, on constate que la classe médiale est identique à la classe médiane, soit [1 800,2 000[. La valeur de la médiale, notée Ml, est ainsi donnée par :

$$Ml = 1\,800 + \frac{200}{0{,}44} \times [0{,}5 - 0{,}35] = 1\,868{,}18 \tag{1.70}$$

▼ **Tableau 1.11** **Salaire mensuel de 150 salariés, en euros**

Classes	n_i	x_i	f_i	F_i	$n_i x_i$	$\frac{n_i x_i}{\sum n_i x_i}$	$\sum \frac{n_i x_i}{\sum n_i x_i}$
[1 400,1 600[	26	1 500	0,17	0,17	39 000	0,14	0,14
[1 600,1 800[	34	1 700	0,23	0,40	57 800	0,21	0,35
[1 800,2 000[	65	1 900	0,43	0,83	123 500	0,44	0,79
[2 000,2 200[	8	2 100	0,05	0,89	16 800	0,06	0,85
[2 200,2 400[	10	2 300	0,07	0,95	23 000	0,08	0,94
[2 400,2 600[	7	2 500	0,05	1	17 500	0,06	1
Total	**150**		**1**		**277 600**	**1**	

Les valeurs de la médiale et de la médiane étant très proches, on peut s'attendre à une très faible concentration des salaires. En comparant l'étendue, 2 600 − 1 400 = 1 200, à la différence entre la médiale et la médiane, ΔM = 1 868,18 − 1 846,51 = 21,67, il ressort une très faible valeur de ΔM par rapport à l'étendue, confirmant la faible concentration des salaires. Ce résultat peut également être illustré par le graphique de la courbe de concentration (► figure 1.21) : l'aire située entre la courbe et la première bissectrice est très proche de zéro, confirmant la très faible concentration des salaires au sein de l'entreprise considérée.



▲ **Figure 1.21** **Courbe de concentration, salaire mensuel en euros des salariés de l'entreprise TextpluS**

Il est possible de calculer un indice de concentration, appelé **indice de Gini** et noté G, égal au double de l'aire de concentration. Il s'agit d'un nombre sans dimension, compris entre 0 et 1 :

- Si $G = 0$, la concentration est nulle.
- Si G est proche de 0, la concentration est faible.
- Si G est proche de 1, la concentration est forte.

Les points clés

- Les tableaux et graphiques statistiques permettent de rassembler et synthétiser visuellement l'information contenue dans les données étudiées.
- Les caractéristiques de tendance centrale, comme la moyenne, le mode ou la médiane, sont des indicateurs de l'ordre de grandeur des données.
- La variance et l'écart-type permettent de quantifier la dispersion des données analysées.
- Les indicateurs de concentration permettent de mesurer des inégalités.

POUR ALLER PLUS LOIN

Les différents types de moyennes

Bien qu'étant la plus utilisée, la moyenne arithmétique n'est qu'un cas particulier de la notion de moyenne. Il existe d'autres types de moyennes, comme la moyenne quadratique, la moyenne géométrique ou encore la moyenne harmonique.

– La **moyenne quadratique** (pondérée) est donnée par :

$$Q = \sqrt{\frac{1}{N}\sum_{i=1}^{k} n_i x_i^2} \qquad (1.71)$$

et est surtout utilisée pour calculer des moyennes d'écarts à une tendance centrale (évitant ainsi le cas de valeurs négatives grâce à l'élévation au carré).

– La **moyenne géométrique** (pondérée) est donnée par :

$$G = \sqrt[N]{x_1^{n_1} \times x_2^{n_2} \times \ldots \times x_k^{n_k}} \qquad (1.72)$$

et est souvent utilisée pour calculer des taux de variation (ou d'accroissement) moyens ou des moyennes de coefficients multiplicateurs.

– La **moyenne harmonique** (pondérée) est donnée par :

$$H = \frac{N}{\sum_{i=1}^{k} \frac{n_i}{x_i}} \qquad (1.73)$$

et est utilisée pour calculer des moyennes de pourcentages ou de rapports, notamment des vitesses et des durées moyennes.

Notons que les différentes moyennes vérifient les inégalités suivantes : $H < G < \bar{x} < Q$.

ÉVALUATION

► Corrigés sur www.dunod.com

QCM

Pour chacune des questions suivantes, indiquer si les affirmations sont vraies ou fausses (il peut y avoir plusieurs réponses vraies pour chaque question).

1 Variables discrètes et variables continues

a. Une variable discrète ne prend que des valeurs positives.

b. Une variable continue est groupée en classes.

c. Le chiffre d'affaires d'une entreprise est une variable discrète.

d. Le nombre de salariés d'une entreprise est une variable discrète.

e. L'âge et la taille sont des variables continues.

2 Caractères et modalités

a. Tout caractère peut avoir une infinité de modalités.

b. Un même individu peut appartenir simultanément à deux (ou plus) modalités.

c. La taille est un caractère quantitatif, de même que l'état matrimonial.

d. Un caractère quantitatif est tel que les modalités qui lui sont associées sont mesurables.

e. Le département de naissance ainsi que la nationalité des individus sont des caractères qualitatifs.

3 Graphiques et centre de classe

a. Un histogramme permet de représenter graphiquement une variable continue.

b. Une variable discrète peut être représentée sous la forme d'un diagramme en bâtons ou d'un diagramme en secteurs.

c. La fonction de répartition est la courbe des fréquences.

d. Lorsque les amplitudes de classes sont différentes, il n'est pas nécessaire de les corriger pour représenter l'histogramme de la distribution correspondante.

e. Le centre de classe peut être calculé par la formule $x_i = \text{extrémité inférieure} + \dfrac{\text{amplitude}}{2}$.

4 Mode

a. Le mode correspond à la valeur de la série qui partage la population en deux sous-ensembles d'effectifs égaux.

b. Le mode est la valeur la plus élevée de la série étudiée.

c. Le mode est égal à la somme des observations de la série divisée par le nombre d'observations.

d. Le mode est une caractéristique de dispersion.

e. Le mode est la valeur de la variable qui correspond à l'effectif le plus élevé.

5 On considère les notes suivantes (sur 20) obtenues par 7 étudiants à l'examen de microéconomie en première année : 18, 15, 8, 12, 8, 15, 4.

a. La moyenne est égale à $\dfrac{18 + 15 + 8 + 12 + 4}{7}$, soit 8,14/20.

b. La note médiane est égale à 12/20.

c. La distribution est unimodale, le mode étant égal à 8/20.

d. L'étendue est égale à 12.

e. Le moment simple d'ordre 1 est égal à 11,42.

Exercice

6 Caractéristiques d'une distribution

On considère un échantillon de 166 agences de location de voitures dans trois régions du Sud de la France. Le tableau 1.12 donne le nombre de voitures louées par jour, avec les effectifs (nombre d'agences de location) correspondants.

▼ Tableau 1.12 Location de voitures

Classes (nombre de voitures louées)	Effectifs n_i
[0,10[	2
[10,20[	19
[20,30[	28
[30,40[	54
[40,50[	31
[50,60[	21
[60,70[	11
Total	**166**

1. La variable étudiée « Nombre de voitures louées » est-elle discrète ou continue ?
2. Calculer les centres de classes, les fréquences et les fréquences cumulées.
3. Quel type de graphique peut-on utiliser pour représenter la série étudiée ?
4. Calculer le mode.
5. Quel est le nombre moyen de véhicules loués par jour ?
6. Calculer la médiane.
7. Déterminer la valeur de l'écart-type de l'échantillon considéré.

Sujet d'examen

7 Université Paris Ouest, extrait

On considère une entreprise dont la répartition des salariés par tranche de salaire (en euros) est reportée dans le tableau 1.13.

1. Quel est l'effectif total de l'entreprise ?
2. En rajoutant autant de colonnes que nécessaire dans le tableau 1.13, calculer :
   – les centres de classes (x_i) ;
   – les amplitudes de classes (a_i) ;
   – les fréquences (f_i) ;
   – les fréquences cumulées (F_i) ;
   – la masse salariale ($n_i x_i$) ;
   – le rapport $\frac{n_i x_i}{\sum_{i=1}^{5} n_i x_i}$;
   – le rapport $\sum_{i=1}^{5} \frac{n_i x_i}{\sum_{i=1}^{5} n_i x_i}$.
3. Calculer le salaire moyen au sein de l'entreprise considérée.
4. Déterminer la classe modale et la valeur du mode.
5. Déterminer l'étendue de la variable étudiée.
6. Déterminer la classe médiane et calculer la valeur de la médiane.
7. Au regard des valeurs prises par les caractéristiques de tendance centrale précédemment calculées, que peut-on en déduire quant à la forme de la distribution des salaires au sein de l'entreprise considérée ?
8. Déterminer la classe médiale et calculer la valeur de la médiale.
9. Calculer le rapport :

$$\frac{\text{Médiale} - \text{Médiane}}{\text{Étendue}} \qquad (1.74)$$

Que peut-on en conclure quant à la concentration des salaires dans cette entreprise ? Ce résultat était-il prévisible ? Pourquoi ?

▼ Tableau 1.13 Répartition des salariés par tranche de salaire

Numéro de classe i	Classes de salaires	Effectifs (n_i)
1	[1 200,1 400[	3
2	[1 400,1 600[	6
3	[1 600,1 800[	182
4	[1 800,2 000[	5
5	[2 000,2 200[	4

Chapitre 2

Si l'on étudie la répartition des salariés au sein d'une entreprise non plus seulement selon leur niveau de salaire, mais également selon l'âge, il s'agit d'une **distribution à deux caractères** : le salaire et l'âge. Si l'on analyse la répartition des salariés de cette même entreprise en fonction simultanément du niveau de salaire, de l'âge et de la catégorie socio-professionnelle, il s'agit d'une distribution à trois caractères, et ainsi de suite. Nous nous limiterons ici à l'étude des distributions à deux caractères[1], sachant que l'exposé peut être aisément généralisé aux distributions à plus de deux caractères.

LES GRANDS AUTEURS



Karl Pearson (1857-1936)

Mathématicien et statisticien britannique, **Karl Pearson** est fréquemment considéré comme l'un des pères de la statistique moderne. Co-fondateur de la célèbre revue *Biometrika*, il est en particulier connu pour ses travaux sur la notion de **corrélation**.
Concept issu de la biologie et introduit en statistique par Francis Galton, la corrélation consiste à étudier la liaison existant entre deux ou plusieurs variables. La mesure de l'intensité de ce lien a été formalisée par Karl Pearson en 1896 au travers du **coefficient de corrélation linéaire**, défini comme le rapport de la covariance entre deux variables sur le produit de leurs écarts-types.
Proche collaborateur de Pearson, George Udny Yule (1871-1951), célèbre statisticien écossais, publia également de nombreux articles sur les notions de corrélation et de dépendance statistique, mais en s'écartant de Pearson quant à l'interprétation et aux hypothèses retenues. ■

1 Comme dans le chapitre 1, nous étudions ici des variables statistiques et non pas des variables aléatoires (► chapitre 6). Les distributions correspondantes sont donc des distributions *empiriques*, de même que leurs caractéristiques associées (fréquences *empiriques*, moyennes *empiriques*, variances *empiriques*, covariance *empirique*, etc.). Toutefois, afin de ne pas alourdir la présentation, le terme « empirique » sera généralement omis dans la suite du chapitre.

Distributions à deux caractères

Plan

Pré-requis

- **Connaître** l'ensemble des concepts étudiés dans le chapitre 1.
- **Savoir calculer** les caractéristiques d'une distribution empirique présentées au chapitre 1 (notamment la moyenne, la variance et l'écart-type).

Objectifs

- **Construire** un tableau statistique à double entrée permettant de synthétiser l'ensemble de l'information pertinente et nécessaire à l'analyse du phénomène considéré.
- **Croiser** l'information afin d'étudier la distribution des effectifs de chaque modalité d'un caractère suivant les modalités de l'autre caractère.
- **Analyser et quantifier** la relation entre deux variables.
- **Évaluer** l'intensité de la liaison entre deux variables.

Dans la mesure où deux variables seront considérées simultanément, par exemple le salaire et l'âge ou le salaire et la catégorie-socio-professionnelle, il sera possible de s'interroger sur l'existence de liens ou de relations entre les données. Ainsi, le salaire augmente-t-il avec l'âge ? Quel est le degré de dépendance du niveau de salaire à la catégorie socio-professionnelle à laquelle appartiennent les salariés ? À niveau de qualification égal, le salaire est-il différent selon la région dans laquelle exerce le salarié ? Le taux de réussite des étudiants à l'université est-il fonction de la catégorie socio-professionnelle de leurs parents ? La rémunération des étudiants diplômés à l'issue de la deuxième année de master est-elle fonction de la discipline principale de leur cursus universitaire ?

Répondre à ces diverses questions nécessite de déterminer si les variables sont liées entre elles. Cela nous conduira à l'analyse de régression et l'étude de la corrélation.

1 Tableaux statistiques à deux dimensions et représentations graphiques

1.1 Un exemple introductif

Reprenons l'étude de l'origine sociale des étudiants à l'université initiée au premier chapitre en complétant l'analyse par l'ajout de diverses informations. Considérons ainsi le tableau 2.1 donnant la répartition des étudiants français dans les principales filières universitaires (y compris IUT) selon leur origine sociale en 2011-2012.

Le tableau 2.1 est un **tableau à double entrée**, encore appelé tableau à deux dimensions ou tableau de contingence : les lignes donnent l'origine sociale (8 modalités) et les colonnes sont relatives à la filière de l'étudiant (6 modalités). Les ligne et colonne intitulées « Total » sont appelées **marges**. Les valeurs figurant dans les cases sont les effectifs et deux lectures du tableau peuvent être réalisées. À titre d'exemple, sur un total de 171 061 étudiants en droit, 15 192 sont issus du milieu ouvrier. De même, sur un total de 123 347 étudiants issus du milieu ouvrier, 15 192 sont inscrits dans la filière « droit ». Plus généralement, à partir des marges du tableau, il est possible de définir deux types de distributions (empiriques) :

- Si l'on associe la ligne « Total » (marge horizontale) et la première ligne du tableau donnant les différentes filières, on obtient la distribution des 1 187 763 étudiants selon la filière suivie. Par exemple, sur les 1 187 763 étudiants considérés, 140 205 sont inscrits en économie. On parle de **distribution marginale** : il s'agit de la distribution empirique marginale du caractère « filière ». Réciproquement, si l'on associe la colonne « Total » (marge verticale) et la première colonne du tableau donnant l'origine sociale, on obtient la distribution des 1 187 763 étudiants selon leur origine sociale. Ainsi, sur les 1 187 763 étudiants considérés, 21 258 sont issus du milieu agricole. Il s'agit ici de la distribution marginale des étudiants selon leur origine sociale.

- Le calcul des pourcentages en ligne et en colonne nous permet de croiser l'information contenue dans les deux caractères, c'est-à-dire d'étudier la distribution des étudiants (i) selon la filière pour chaque origine sociale et (ii) selon leur origine sociale dans chaque filière.

▼ **Tableau 2.1 Origine sociale des étudiants à l'université en 2011-2012**

Filière / Origine sociale	Droit	Économie	Lettres	Sciences	Santé	IUT	Total
Agriculteurs	2 543	2 665	5 508	4 788	2 999	2 755	21 258
Artisans, commerçants, chefs d'entreprise	15 384	12 029	22 819	16 379	11 784	9 549	87 944
Professions libérales, cadres supérieurs	60 732	34 867	91 046	72 033	74 984	29 620	363 282
Professions intermédiaires	18 008	14 666	47 672	33 135	20 887	16 862	151 230
Employés	19 984	17 186	47 543	30 359	14 173	15 323	144 568
Ouvriers	15 192	16 601	39 061	27 097	10 406	14 990	123 347
Retraités, inactifs	24 368	21 506	58 154	27 257	16 119	9 415	156 819
Non renseigné	14 850	20 685	45 079	23 579	30 201	4 921	139 315
Total	**171 061**	**140 205**	**356 882**	**234 627**	**181 553**	**103 435**	**1 187 763**

Source : Ministère de l'Enseignement Supérieur et de la Recherche, MESR (DGESIP-DGRI-SIES).

Le tableau 2.2, basé sur le calcul des pourcentages en ligne, donne ainsi pour chacune des 8 modalités du caractère « origine sociale » la distribution des étudiants selon la filière. En d'autres termes, il s'agit de la distribution des étudiants entre les différentes filières *conditionnellement à* (c'est-à-dire sachant) leur origine sociale. On constate que 31,67 % des étudiants provenant du milieu ouvrier sont inscrits en lettres, alors que seuls 8,44 % d'entre eux sont inscrits dans la filière « santé ». On parle ici de **distribution conditionnelle** : il s'agit de la distribution empirique conditionnelle des étudiants selon la filière pour chaque origine sociale.

De façon similaire, le tableau 2.3, reportant les valeurs des pourcentages en colonne, donne la distribution conditionnelle des étudiants selon leur origine sociale au sein de chaque filière : 41,3 % des étudiants inscrits en santé sont issus de professions libérales et cadres supérieurs, alors que seuls 1,65 % d'entre eux sont issus du milieu agricole.

Cet exemple met en évidence l'intérêt des tableaux à deux dimensions puisqu'ils offrent la possibilité d'étudier la distribution des effectifs de chaque modalité d'un caractère suivant les modalités de l'autre. Les sections suivantes généralisent les notions abordées dans le cadre de cet exemple.

▼ Tableau 2.2 Distribution conditionnelle des étudiants selon la filière pour chaque origine sociale

Origine sociale \ Filière	Droit	Économie	Lettres	Sciences	Santé	IUT	Total
Agriculteurs	11,96	12,54	25,91	22,52	14,11	12,96	**100**
Artisans, commerçants, chefs d'entreprise	17,49	13,68	25,95	18,62	13,40	10,86	**100**
Professions libérales, cadres supérieurs	16,72	9,60	25,06	19,83	20,64	8,15	**100**
Professions intermédiaires	11,91	9,70	31,52	21,91	13,81	11,15	**100**
Employés	13,82	11,89	32,89	21,00	9,80	10,60	**100**
Ouvriers	12,32	13,46	31,67	21,97	8,44	12,15	**100**
Retraités, inactifs	15,54	13,71	37,08	17,38	10,28	6,00	**100**
Non renseigné	10,66	14,85	32,36	16,92	21,68	3,53	**100**
Total	14,40	11,80	30,05	19,75	15,29	8,71	**100**

▼ Tableau 2.3 Distribution conditionnelle des étudiants selon l'origine sociale dans chaque filière

Origine sociale \ Filière	Droit	Économie	Lettres	Sciences	Santé	IUT	Total
Agriculteurs	1,49	1,90	1,54	2,04	1,65	2,66	1,79
Artisans, commerçants, chefs d'entreprise	8,99	8,58	6,39	6,98	6,49	9,23	7,40
Professions libérales, cadres supérieurs	35,50	24,87	25,51	30,70	41,30	28,64	30,59
Professions intermédiaires	10,53	10,46	13,36	14,12	11,50	16,30	12,73
Employés	11,68	12,26	13,32	12,94	7,81	14,81	12,17
Ouvriers	8,88	11,84	10,95	11,55	5,73	14,49	10,38
Retraités, inactifs	14,25	15,34	16,30	11,62	8,88	9,10	13,20
Non renseigné	8,68	14,75	12,63	10,05	16,63	4,76	11,73
Total	**100**	**100**	**100**	**100**	**100**	**100**	**100**

1.2 Forme générale des tableaux à deux dimensions

Considérons un échantillon composé de N individus, chacun d'entre eux étant doté de deux caractères. Notons ces caractères (ou variables) x et y comportant respectivement r et s modalités : $x_1, x_2, \ldots, x_i, \ldots, x_r$ et $y_1, y_2, \ldots, y_j, \ldots, y_s$. Afin de construire le tableau statistique correspondant, il convient de comptabiliser les individus ayant simultanément les modalités x_i et y_j pour $i = 1,\ldots,r$ et $j = 1,\ldots,s$. Le nombre obtenu, noté n_{ij}, correspond à l'effectif des individus caractérisés simultanément par les modalités x_i et y_j. On reporte alors les valeurs des effectifs dans un tableau à deux dimensions (▸ tableau 2.4), les modalités de x figurant en ligne, celles de y en colonne.

▼ Tableau 2.4 Tableau statistique à deux dimensions

x_i \ y_j	y_1	y_2	...	y_j	...	y_s	Total
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1s}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2s}	$n_{2\bullet}$
⋮	⋮	⋮		⋮		⋮	⋮
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{is}	$n_{i\bullet}$
⋮	⋮	⋮		⋮		⋮	⋮
x_r	n_{r1}	n_{r2}	...	n_{rj}	...	n_{rs}	$n_{r\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet s}$	$n_{\bullet\bullet}$

1.2.1 Effectifs

La forme générale d'un tableau statistique à deux dimensions telle que celle présentée dans le tableau 2.4 appelle quelques précisions concernant les notations utilisées. Dans les marges (ligne et colonne « Total »), l'indice sur lequel est opéré la sommation est remplacé par un point. Ainsi, si l'on considère la ligne i du tableau, $n_{i\bullet}$ correspond à somme des effectifs de cette ligne (appelés **effectifs marginaux** de x) :

$$n_{i\bullet} = n_{i1} + n_{i2} + \ldots + n_{ij} + \ldots + n_{is} = \sum_{j=1}^{s} n_{ij} \tag{2.1}$$

De même, $n_{\bullet j}$ correspond à la somme des effectifs de la colonne j (c'est-à-dire aux effectifs marginaux de y), la somme étant effectuée sur l'indice i :

$$n_{\bullet j} = n_{1j} + n_{2j} + \ldots + n_{ij} + \ldots + n_{rj} = \sum_{i=1}^{r} n_{ij} \tag{2.2}$$

Si l'on effectue la somme ligne par ligne de l'ensemble des lignes, on obtient :

$$n_{\bullet\bullet} = n_{1\bullet} + n_{2\bullet} + \ldots + n_{i\bullet} + \ldots + n_{r\bullet} = \sum_{i=1}^{r} \sum_{j=1}^{s} n_{ij} \tag{2.3}$$

soit finalement :

$$n_{\bullet\bullet} = \sum_{i=1}^{r} n_{i\bullet} = N \tag{2.4}$$

Par un calcul similaire et en effectuant la somme colonne par colonne de l'ensemble des colonnes, il vient :

$$n_{\bullet\bullet} = \sum_{j=1}^{s} \sum_{i=1}^{r} n_{ij} = \sum_{j=1}^{s} n_{\bullet j} = N \tag{2.5}$$

En résumé, $n_{\bullet\bullet}$ correspond donc à l'**effectif total** de la population considérée N :

$$n_{\bullet\bullet} = \sum_{i=1}^{r} n_{i\bullet} = \sum_{j=1}^{s} n_{\bullet j} = N \tag{2.6}$$

1.2.2 Fréquences

Tout comme pour les distributions à un caractère, il est possible de calculer des fréquences (empiriques) dans le cas des distributions à deux caractères et de reporter leurs valeurs dans les tableaux. C'est ce que nous avons effectué dans le cadre de notre exemple au sein des tableaux 2.2 et 2.3. Les fréquences sont désormais associées, non plus à une seule modalité, mais au couple de valeurs (x_i, y_j). La fréquence f_{ij} correspond ainsi à la proportion d'individus présentant simultanément les modalités x_i et y_j et est donnée par :

$$f_{ij} = \frac{n_{ij}}{N} = \frac{n_{ij}}{n_{\bullet\bullet}} \tag{2.7}$$

Propriétés

En utilisant les mêmes notations que pour les effectifs, on déduit les propriétés suivantes pour les fréquences :

– Le total des fréquences de la ligne i est donné par :

$$f_{i\bullet} = \sum_{j=1}^{s} f_{ij} = \sum_{j=1}^{s} \frac{n_{ij}}{N} = \frac{n_{i\bullet}}{N} \tag{2.8}$$

– Le total des fréquences de la colonne j est donné par :

$$f_{\bullet j} = \sum_{i=1}^{r} f_{ij} = \sum_{i=1}^{r} \frac{n_{ij}}{N} = \frac{n_{\bullet j}}{N} \tag{2.9}$$

– La somme des fréquences est égale à 1 :

$$\sum_{i=1}^{r}\sum_{j=1}^{s} f_{ij} = \sum_{i=1}^{r} f_{i\bullet} = \sum_{j=1}^{s} f_{\bullet j} = 1 \tag{2.10}$$

On constate ainsi que les fréquences $f_{i\bullet}$ et $f_{\bullet j}$ sont définies par le rapport entre les effectifs marginaux et l'effectif total, elles sont appelées **fréquences marginales** : $f_{i\bullet}$ (respectivement $f_{\bullet j}$) est la fréquence empirique marginale de la modalité x_i (resp. y_j) du caractère x (resp. y). À titre d'exemple, en reprenant les données du tableau 2.1, il apparaît que :

– $f_{1\bullet} = \frac{n_{1\bullet}}{n_{\bullet\bullet}} = \frac{21\,258}{1\,187\,763} = 0{,}018$: 1,8 % des étudiants considérés sont issus du milieu agricole.

– $f_{\bullet 4} = \frac{n_{\bullet 4}}{n_{\bullet\bullet}} = \frac{234\,627}{1\,187\,763} = 0{,}1\,975$: 19,75 % des étudiants considérés sont inscrits dans la filière « sciences ».

Les fréquences marginales se rapportent ainsi aux distributions à une dimension puisque la distribution marginale suivant le caractère x ne fait pas intervenir le caractère y ; de même pour la distribution marginale suivant le caractère y qui ne tient pas compte du caractère x.

Si l'on considère à présent simultanément l'information contenue dans les deux caractères x et y, on peut calculer des **fréquences conditionnelles**. La fréquence empirique

conditionnelle de x selon y, notée $f_{i/j}$, correspond à la proportion d'individus présentant la modalité x_i parmi les individus présentant uniquement la modalité y_j :

$$f_{i/j} = \frac{n_{ij}}{n_{\bullet j}} \tag{2.11}$$

Si l'on reprend le tableau 2.3, on constate que 14,49 % des étudiants inscrits en IUT proviennent du milieu ouvrier. De façon équivalente, on peut dire que sur les 103 435 étudiants inscrits en IUT, 14 990 sont issus du milieu agricole (▸ tableau 2.1). De façon générale, les différentes fréquences conditionnelles pour une même modalité x_i du caractère x donnent la distribution conditionnelle de x selon y.

On définit de même la fréquence conditionnelle de y selon x, notée $f_{j/i}$, comme la proportion d'individus présentant la modalité y_j parmi les individus présentant uniquement la modalité x_i :

$$f_{j/i} = \frac{n_{ij}}{n_{i\bullet}} \tag{2.12}$$

En reprenant les données des tableaux 2.1 et 2.2, il apparaît que parmi les 87 944 étudiants issus du milieu « artisans, commerçants, chefs d'entreprise », 22 819 sont inscrits dans la filière « lettres ». De façon similaire, 25,95 % des étudiants issus du milieu « artisans, commerçants, chefs d'entreprise » sont inscrits dans la filière « lettres ». De façon générale et comme précédemment, les différentes fréquences conditionnelles pour une même modalité y_j du caractère y donnent la distribution conditionnelle de y selon x.

Propriétés

– De même que pour toutes les fréquences, la somme des fréquences conditionnelles est égale à l'unité :

$$\sum_{i=1}^{r} f_{i/j} = \sum_{i=1}^{r} \frac{n_{ij}}{n_{\bullet j}} = 1 \tag{2.13}$$

et :

$$\sum_{j=1}^{s} f_{j/i} = \sum_{j=1}^{s} \frac{n_{ij}}{n_{i\bullet}} = 1 \tag{2.14}$$

– Les fréquences conditionnelles et marginales sont liées par la relation :

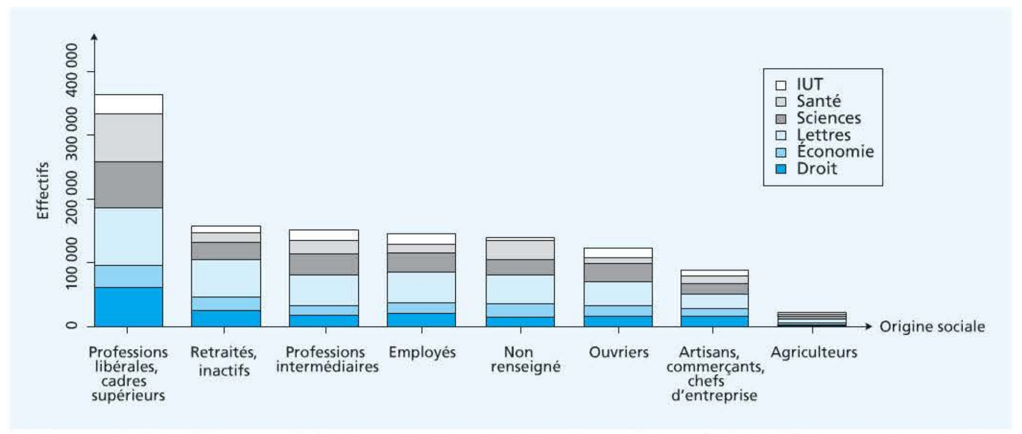
$$f_{ij} = f_{i\bullet} \times f_{j/i} = f_{\bullet j} \times f_{i/j} \tag{2.15}$$

1.3 Représentations graphiques

Comme dans le cas des distributions à un caractère (▸ chapitre 1), le type de représentation graphique adéquat pour les distributions à deux caractères dépend de la nature des caractères étudiés. Les caractères pouvant être qualitatifs, quantitatifs discrets ou quantitatifs continus, trois principaux cas peuvent se présenter[2] :

2 Notons qu'il existe de très nombreux autres types de graphiques que ceux mentionnés ici, comme les diagrammes cartésiens, les cartogrammes, les diagrammes de fréquences... Le lecteur intéressé pourra notamment consulter les ouvrages de Grais (2003) ou Py (2007).

- Les deux caractères étudiés sont qualitatifs. Dans ce cas et comme pour les distributions à un caractère, on peut utiliser les représentations en tuyaux d'orgue ou par secteurs. La figure 2.1 donne ainsi la représentation en tuyaux d'orgue des données reportées dans le tableau 2.1. La distribution marginale selon l'origine sociale est représentée par les hauteurs (et les surfaces) des tuyaux d'orgue. À l'intérieur de ces tuyaux, on représente par des rectangles les effectifs figurant dans le tableau ; la hauteur (et la surface) de chaque rectangle donnant la valeur de l'effectif correspondant.



▲ **Figure 2.1 Origine sociale des étudiants à l'université en 2011-2012, représentation en tuyaux d'orgue**

- Les deux caractères étudiés sont quantitatifs. On utilise alors en général des séries d'histogrammes ou des stéréogrammes, ces derniers étant des histogrammes à trois dimensions (effectifs ou fréquences, modalités du caractère x, modalités du caractère y).
- Un caractère est qualitatif, l'autre est quantitatif. On utilise des représentations en tuyaux d'orgue ou des diagrammes en bâtons.

2 Caractéristiques des distributions à deux caractères

Dans la mesure où il existe deux types de distributions, marginales et conditionnelles, il est possible de déterminer des caractéristiques pour chacune de ces deux distributions.

2.1 Caractéristiques des distributions marginales

Les caractéristiques marginales du caractère x se déterminent à partir de la distribution marginale de ce même caractère, c'est-à-dire à partir des r modalités x_i et des r effec-

tifs marginaux $n_{i\bullet}$, $i = 1,\ldots,r$. De façon similaire, les caractéristiques marginales de y s'obtiennent à partir de la distribution marginale de ce même caractère, c'est-à-dire à partir des s modalités y_j et des s effectifs marginaux $n_{\bullet j}$, $j = 1,\ldots,s$. On peut ainsi définir :

- La moyenne (empirique) marginale de x :

$$\bar{x} = \frac{1}{n_{\bullet\bullet}} \sum_{i=1}^{r} n_{i\bullet} x_i = \sum_{i=1}^{r} f_{i\bullet} x_i \tag{2.16}$$

- La moyenne (empirique) marginale de y :

$$\bar{y} = \frac{1}{n_{\bullet\bullet}} \sum_{j=1}^{s} n_{\bullet j} y_j = \sum_{j=1}^{s} f_{\bullet j} y_j \tag{2.17}$$

- La variance (empirique) marginale de x :

$$V(x) = \frac{1}{n_{\bullet\bullet}} \sum_{i=1}^{r} n_{i\bullet}(x_i - \bar{x})^2 = \frac{1}{n_{\bullet\bullet}} \sum_{i=1}^{r} n_{i\bullet} x_i^2 - \bar{x}^2 \tag{2.18}$$

- La variance (empirique) marginale de y :

$$V(y) = \frac{1}{n_{\bullet\bullet}} \sum_{j=1}^{s} n_{\bullet j}(y_j - \bar{y})^2 = \frac{1}{n_{\bullet\bullet}} \sum_{j=1}^{s} n_{\bullet j} y_j^2 - \bar{y}^2 \tag{2.19}$$

2.2 Caractéristiques des distributions conditionnelles

Contrairement aux distributions marginales, les caractéristiques des distributions conditionnelles tiennent compte des modalités des deux caractères. De façon pratique, on se donne une modalité d'un caractère, par exemple y_j, et l'on étudie la façon dont l'effectif de cette modalité se distribue entre l'ensemble des modalités de l'autre caractère, x. Les caractéristiques conditionnelles de x selon y sont ainsi déterminées à partir des s distributions conditionnelles de x selon y, c'est-à-dire à partir des r modalités de x et des s colonnes d'effectifs du tableau à deux dimensions associées à ces r modalités. De façon similaire, les caractéristiques conditionnelles de y selon x sont obtenues à partir des r distributions conditionnelles de y selon x, c'est-à-dire à partir des s modalités de y et des r lignes d'effectifs du tableau à deux dimensions associées à ces s modalités. On peut dès lors définir :

– Les moyennes (empiriques) conditionnelles de x selon y :

$$\bar{x}_j = \frac{1}{n_{\bullet j}} \sum_{i=1}^{r} n_{ij} x_i = \sum_{i=1}^{r} f_{i/j} x_i \tag{2.20}$$

– Les moyennes (empiriques) conditionnelles de y selon x :

$$\bar{y}_i = \frac{1}{n_{i\bullet}} \sum_{j=1}^{s} n_{ij} y_j = \sum_{j=1}^{s} f_{j/i} y_j \tag{2.21}$$

– Les variances (empiriques) conditionnelles de x selon y :

$$V_j(x) = \frac{1}{n_{\bullet j}} \sum_{i=1}^{r} n_{ij}(x_i - \bar{x}_j)^2 = \frac{1}{n_{\bullet j}} \sum_{i=1}^{r} n_{ij}x_i^2 - \bar{x}_j{}^2 \qquad (2.22)$$

– Les variances (empiriques) conditionnelles de y selon x :

$$V_i(y) = \frac{1}{n_{i\bullet}} \sum_{j=1}^{s} n_{ij}(y_j - \bar{y}_i)^2 = \frac{1}{n_{i\bullet}} \sum_{j=1}^{s} n_{ij}y_j^2 - \bar{y}_i{}^2 \qquad (2.23)$$

Propriétés

Il est possible de mettre en évidence des relations entre (i) les moyennes marginale et conditionnelle et (ii) les variances marginale et conditionnelle.

– Relations entre les moyennes :

$$\bar{x} = \frac{1}{n_{\bullet\bullet}} \sum_{j=1}^{s} n_{\bullet j}\bar{x}_j \qquad (2.24)$$

$$\bar{y} = \frac{1}{n_{\bullet\bullet}} \sum_{i=1}^{r} n_{i\bullet}\bar{y}_i \qquad (2.25)$$

– Relations entre les variances :

$$V(x) = \underbrace{\frac{1}{n_{\bullet\bullet}} \sum_{j=1}^{s} n_{\bullet j}(\bar{x}_j - \bar{x})^2}_{\text{variance des moyennes conditionnelles } \bar{x}_j} + \underbrace{\frac{1}{n_{\bullet\bullet}} \sum_{j=1}^{s} V_j(x)n_{\bullet j}}_{\text{moyenne des variances conditionnelles } V_j(x)} \qquad (2.26)$$

$$V(y) = \underbrace{\frac{1}{n_{\bullet\bullet}} \sum_{i=1}^{r} n_{i\bullet}(\bar{y}_i - \bar{y})^2}_{\text{variance des moyennes conditionnelles } \bar{y}_i} + \underbrace{\frac{1}{n_{\bullet\bullet}} \sum_{i=1}^{r} V_i(y)n_{i\bullet}}_{\text{moyenne des variances conditionnelles } V_i(y)} \qquad (2.27)$$

soit encore :

$$V(x) = V(\bar{x}_j) + \overline{V_j(x)} \qquad (2.28)$$

et :

$$V(y) = V(\bar{y}_i) + \overline{V_i(y)} \qquad (2.29)$$

La variance marginale est ainsi égale à la somme de la variance des moyennes conditionnelles et de la moyenne des variances conditionnelles. La dispersion de la distribution marginale est donc fonction de la dispersion entre les moyennes conditionnelles et de la dispersion de chacune des distributions conditionnelles autour de leur moyenne[3].

3 Dans le cas où l'on étudie une population (par exemple des pays) composée de deux (ou plusieurs) sous-populations (pays développés et pays en développement), la variance des moyennes est également appelée *variance inter-population* et la moyenne des variances, *variance intra-population*.

2.3 Covariance et notion de dépendance

Lorsque l'on étudie des distributions à deux caractères, on cherche généralement à quantifier le lien entre ceux-ci. Ce point fera l'objet d'une étude approfondie dans la section suivante, nous présentons ici un indicateur fréquemment calculé : la covariance. Ainsi que nous le verrons par la suite, cet indicateur servira dans l'étude de la corrélation entre deux variables.

Définition 2.1

La **covariance** (empirique) entre les variables x et y, notée $Cov(x,y)$, est donnée par :

$$Cov(x,y) = \frac{1}{n_{\bullet\bullet}} \sum_{i=1}^{r} \sum_{j=1}^{s} n_{ij}(x_i - \bar{x})(y_j - \bar{y}) \tag{2.30}$$

soit encore, sous forme développée :

$$Cov(x,y) = \frac{1}{n_{\bullet\bullet}} \sum_{i=1}^{r} \sum_{j=1}^{s} n_{ij}x_i y_j - \bar{x}\bar{y} \tag{2.31}$$

Comme nous l'étudierons dans la section suivante, l'analyse de régression permettra de quantifier le lien entre deux variables.

FOCUS

La notion de dépendance

Trois types de liens ou de liaisons peuvent être mis en évidence à partir de l'étude simultanée de deux caractères :

- Lorsque les deux variables x et y ne présentent aucun lien entre elles, c'est-à-dire si les variations d'une variable ne s'accompagnent pas de variations de l'autre variable, on dit que x et y sont **indépendantes**. Dans ce cas, les fréquences conditionnelles sont égales aux fréquences marginales :

$$f_{i/j} = f_{i\bullet} \quad \text{et} \quad f_{j/i} = f_{\bullet j} \tag{2.32}$$

et les moyennes marginales et conditionnelles sont identiques pour chacune des deux variables.

- Si à chaque valeur de x correspond une et une seule valeur de y parfaitement déterminée et réciproquement, on dit que les variables x et y sont totalement (ou parfaitement) dépendantes ou encore qu'il existe une **liaison fonctionnelle** entre x et y. Dans ce cas, les moyennes conditionnelles sont égales aux valeurs des variables, soit :

$$\bar{x}_j = x_i \quad \text{et} \quad \bar{y}_i = y_j \tag{2.33}$$

- En pratique, dans la plupart des cas, les phénomènes étudiés se situent entre ces deux cas extrêmes : les variables ne sont pas indépendantes, mais ne sont pas non plus parfaitement dépendantes. On parle alors de **corrélation**. Lorsque les deux variables x et y évoluent dans le même sens, on parle de corrélation positive. Lorsque les deux variables évoluent en sens contraire (l'une augmente quand l'autre diminue), on parle de corrélation négative. Ce cas est étudié en détail dans la section qui suit *via* l'analyse de régression.

3 Liens entre deux variables : régression et corrélation

Ainsi que nous l'avons précédemment mentionné, le recours aux distributions à deux caractères rend possible l'étude du lien entre les deux variables considérées. Notons, pour toute la suite du chapitre, x et y ces deux variables et supposons que celles-ci soient observées sur les N individus statistiques i étudiés, $i = 1,...,N$. N désigne donc le nombre d'observations. La variable x prend en conséquence les valeurs $x_1,...,x_i,...,x_N$ et la variable y les valeurs $y_1,...,y_i,...,y_N$. Pour chaque individu i, on observe ainsi deux mesures.

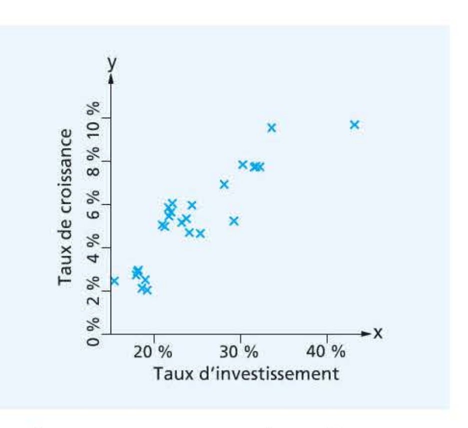
Considérons, à titre d'exemple, un échantillon de 26 pays i ($i = 1,...,N$ avec $N = 26$) pour lesquels on observe les deux variables suivantes pour l'année 2010 : le taux d'investissement en pourcentage du PIB, noté x_i, et le taux de croissance économique (exprimé en pourcentage), noté y_i, avec $i = 1,...,26$. Les données sont reportées dans le tableau 2.5.

▼ **Tableau 2.5 Taux d'investissement en pourcentage du PIB et taux de croissance économique en 2010 sur un échantillon de 26 régions ou pays**

x_i	y_i	x_i	y_i
15,39	2,45	23,18	5,15
17,91	2,72	23,73	5,32
18,07	2,87	24,07	4,68
18,15	2,95	24,37	5,95
18,58	2,11	25,34	4,64
18,97	2,50	28,08	6,91
19,20	2,03	29,20	5,22
20,92	5,04	30,24	7,82
21,22	4,96	31,57	7,70
21,64	5,84	31,76	7,74
21,72	5,44	32,22	7,72
21,93	5,62	33,57	9,53
22,11	6,03	43,15	9,67

Source : World Bank, Word Development Indicators (WDI).

Au niveau économique, il est raisonnable de penser que les variables sont liées, l'investissement étant souvent considéré comme un moteur de la croissance. Une façon simple d'appréhender cette relation est de représenter graphiquement le **nuage de points**, c'est-à-dire l'ensemble des couples (x_i,y_i) ainsi que l'illustre la figure 2.2. Les valeurs du taux d'investissement sont reportées en abscisse, celles du taux de croissance en ordonnée. On constate que le nuage de points exhibe une forme allongée, les points semblant relativement peu dispersés dans le plan (x,y). Plus spécifiquement, il apparaît une relation croissante entre les deux variables : lorsque le taux d'investissement augmente, la croissance augmente et réciproquement. On parle dans ce cas de **corrélation** positive entre les variables, corrélation que nous quantifierons dans la suite de l'exposé.



▲ **Figure 2.2 Nuage de points, taux d'investissement et taux de croissance économique**

3.1 L'analyse de régression et le principe des moindres carrés ordinaires

FOCUS

Analyse de régression et économétrie

L'analyse de **régression** renvoie à un ensemble de méthodes visant à analyser la relation existant entre deux ou plusieurs variables. Plus spécifiquement, si l'on considère le cas de deux variables x et y, l'objectif est d'expliquer la variable y (appelée **variable expliquée**) par la variable x (appelée **variable explicative**). Il existe plusieurs types de modèles de régression, le plus connu étant le **modèle de régression linéaire** : $y = ax + b$ où a et b sont des constantes désignant respectivement le coefficient de pente de la droite de régression de y sur x et l'ordonnée à l'origine. À titre d'exemple, un tel modèle peut être utilisé comme représentation de la fonction de consommation keynésienne : $C = cR + c_0$ où C désigne la consommation, R est le revenu, c et c_0 sont des paramètres tels que $0 < c < 1$ et $c_0 > 0$, c désignant la propension marginale à consommer.

Ce modèle de régression linéaire traduit l'existence d'une relation (linéaire) croissante entre la consommation et le revenu (c est positif), l'accroissement de la consommation étant moindre que celui du revenu (c est inférieur à 1), conformément à la loi psychologique fondamentale de Keynes. L'analyse de régression est l'une des méthodes les plus utilisées en statistique et en **économétrie**.

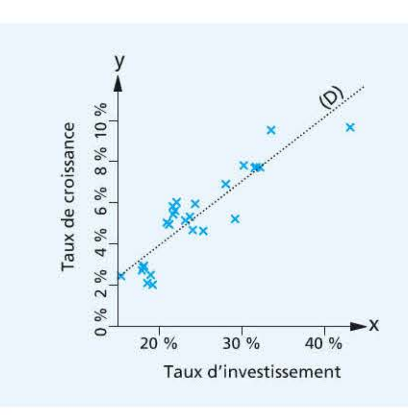
De façon générale, l'économétrie renvoie à la mesure des phénomènes économiques et permet d'analyser, d'estimer et de vérifier, c'est-à-dire de tester, les phénomènes et théories économiques. L'économétrie constitue ainsi une branche de la science économique qui fait appel conjointement à la théorie économique, la statistique, les mathématiques et l'informatique. En tant que discipline, elle est née en 1930 lors de la création de la Société d'économétrie par Ragnar Frisch, Charles Roos et Irving Fisher[4].

3.1.1 Principe général

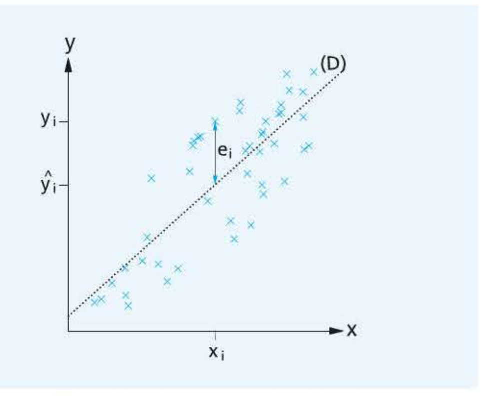
Afin de mettre plus clairement en évidence et de quantifier la relation entre nos deux variables, le taux d'investissement et le taux de croissance économique, il convient de « résumer » le nuage de points, c'est-à-dire de représenter dans le plan (x,y) l'allure générale de la distribution à deux caractères. À cette fin, la méthode la plus utilisée consiste à ajuster le nuage de points par une droite (D), comme cela est reproduit sur la figure 2.3. On parle de **droite de régression** ou de **droite d'ajustement** ou encore de **droite des moindres carrés**. La droite (D) ne passant pas par tous les points du nuage, il existe naturellement des **écarts** entre les points du nuage et les points situés sur cette droite.

POUR ALLER PLUS LOIN
▸ Voir p. 59

4 Le lecteur intéressé par l'économétrie pourra notamment consulter les manuels de Gouriéroux et Monfort (1989), Greene (2005) ou Mignon (2008).



▲ Figure 2.3 Taux d'investissement et taux de croissance économique, ajustement du nuage de points par une droite



▲ Figure 2.4 Ajustement du nuage de points par une droite de régression

Plus généralement, comme reproduit sur la figure 2.4, notons $\hat{y}_i$ l'ordonnée d'un point de la droite (D) dont l'abscisse est x_i et désignons par e_i les écarts entre la valeur observée y_i de y et la valeur $\hat{y}_i$ située sur la droite :

$$e_i = y_i - \hat{y}_i \quad (2.34)$$

L'expression de la droite de régression est alors donnée par :

$$\hat{y}_i = ax_i + b \quad (2.35)$$

où a et b sont des constantes, et le problème consiste à identifier la droite (D) qui ajuste au mieux le nuage de points considéré. En d'autres termes, il s'agit de trouver la droite (D) telle que les écarts e_i soient les plus faibles possibles, c'est-à-dire telle que les valeurs $\hat{y}_i$ situées sur la droite soient les plus proches possibles des valeurs observées y_i.

La **méthode des moindres carrés ordinaires** (MCO) nous permet précisément de répondre à cet objectif puisqu'elle consiste à trouver la droite (D), c'est-à-dire les valeurs des paramètres a et b, telles que la somme des carrés des écarts e_i soit minimale. Notons que l'on cherche à minimiser la somme des carrés des écarts et non pas directement les écarts puisque ceux-ci pouvant prendre des valeurs positives et négatives (et nulles), ils peuvent se compenser de sorte que leur somme – et donc leur moyenne – reste proche de zéro.

3.1.2 Détermination de l'équation de la droite de régression

Afin de déterminer la valeur des paramètres (ou coefficients) a et b de la droite de régression, on applique le principe des MCO consistant à minimiser la somme des carrés des écarts. On cherche ainsi à minimiser l'expression suivante par rapport aux paramètres a et b :

$$\sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} (y_i - ax_i - b)^2 \quad (2.36)$$

Propriété

Les paramètres a et b de la droite de régression sont donnés par :

$$a = \frac{Cov(x,y)}{V(x)} \qquad \text{et} \qquad b = \bar{y} - a\bar{x} \tag{2.37}$$

$V(x)$ désignant la variance de x et $Cov(x,y)$ la covariance entre x et y :

$$V(x) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \bar{x}^2 \tag{2.38}$$

$$Cov(x,y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \bar{x}\bar{y} \tag{2.39}$$

Le paramètre a est la pente de la droite de régression de y sur x, b désignant l'ordonnée à l'origine.

Démonstration

On cherche à minimiser l'expression $S = \sum_{i=1}^{N}(y_i - ax_i - b)^2$ par rapport à a et b. À cette fin, on annule les dérivées partielles de S par rapport à a et b, soit :

$$\frac{\partial S}{\partial a} = -2\sum_{i=1}^{N}(y_i - ax_i - b)x_i = 0 \tag{2.40}$$

$$\frac{\partial S}{\partial b} = -2\sum_{i=1}^{N}(y_i - ax_i - b) = 0 \tag{2.41}$$

On obtient alors un système de deux équations à deux inconnues :

$$\sum_{i=1}^{N} x_i y_i = a\sum_{i=1}^{N} x_i^2 + b\sum_{i=1}^{N} x_i \tag{2.42}$$

$$\sum_{i=1}^{N} y_i = a\sum_{i=1}^{N} x_i + Nb \tag{2.43}$$

En divisant les deux membres de l'équation (2.43) par N, il vient :

$$\bar{y} = a\bar{x} + b \tag{2.44}$$

stipulant que la droite de régression passe par le point moyen de coordonnées $(\bar{x},\bar{y})$. On en déduit la valeur de b :

$$b = \bar{y} - a\bar{x} \tag{2.45}$$

En remplaçant b par son expression dans l'équation (2.42), on obtient :

$$\sum_{i=1}^{N} x_i y_i = a\sum_{i=1}^{N} x_i^2 + (\bar{y} - a\bar{x})\sum_{i=1}^{N} x_i \tag{2.46}$$

Soit encore, en remplaçant $\bar{x}$ et $\bar{y}$ par leurs expressions :

$$\sum_{i=1}^{N} x_i y_i - \frac{1}{N}\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i = a\left[\sum_{i=1}^{N} x_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} x_i\right)^2\right] \tag{2.47}$$

Sachant que la variance de x est donnée par :

$$V(x) = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \bar{x}^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)^2 \tag{2.48}$$

et que la covariance entre x et y est définie par :

$$Cov(x,y) = \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \bar{x}\bar{y} = \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \frac{1}{N}\sum_{i=1}^{N} x_i \frac{1}{N}\sum_{i=1}^{N} y_i \quad (2.49)$$

on en déduit que la relation (2.47) peut s'exprimer comme suit :

$$NCov(x,y) = aNV(x) \quad (2.50)$$

D'où la valeur recherchée pour a :

$$a = \frac{Cov(x,y)}{V(x)} \quad (2.51)$$

■

Si l'on reprend l'exemple de la relation entre le taux d'investissement et le taux de croissance économique (▶ tableau 2.5), on peut calculer les divers indicateurs nécessaires à la détermination de a et b :

– Moyenne du taux de croissance : $\bar{y} = 5{,}33$

– Moyenne du taux d'investissement : $\bar{x} = 24{,}47$

– Covariance entre le taux d'investissement et le taux de croissance : $Cov(x,y) = 12{,}14$

– Variance du taux d'investissement : $V(x) = 39{,}11$

On obtient alors aisément les valeurs des paramètres de la droite de régression :

$$a = \frac{12{,}14}{39{,}11} = 0{,}31 \qquad \text{et} \qquad b = 5{,}33 - 0{,}31 \times 24{,}47 = -2{,}26 \quad (2.52)$$

D'où l'équation de la droite de régression reportée sur la figure 2.3 :

$$\hat{y} = 0{,}31x - 2{,}26 \quad (2.53)$$

On constate ainsi que la valeur de a est positive, illustrant bien l'existence d'une relation positive entre les deux variables : lorsque le taux d'investissement augmente, la croissance augmente également toutes choses égales par ailleurs.

3.2 Mesure du degré de liaison entre deux variables et qualité d'une régression

3.2.1 Coefficient de corrélation linéaire

Le coefficient de corrélation linéaire est un nombre sans dimension permettant de quantifier le degré de liaison (linéaire) entre deux variables. Il s'agit ainsi d'un indicateur du degré de proximité entre les points du nuage et ceux figurant sur la droite de régression.

Définition 2.2

Le **coefficient de corrélation linéaire** $r(x,y)$ entre les variables x et y est donné par :

$$r(x,y) = \frac{Cov(x,y)}{\sigma_x \sigma_y} \tag{2.54}$$

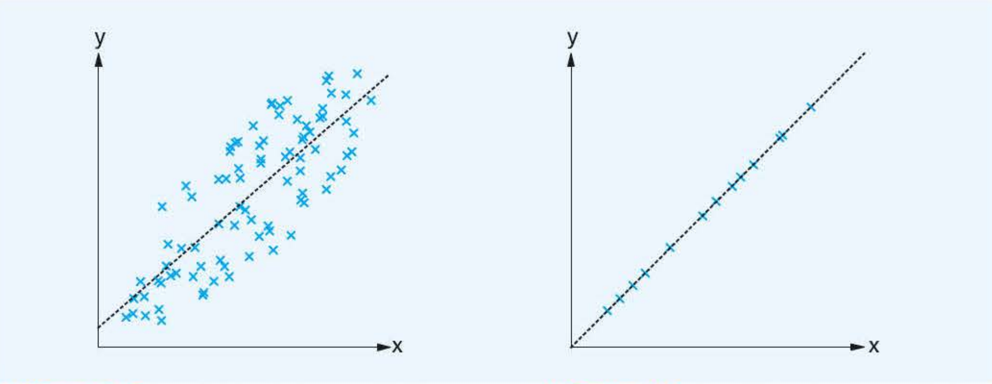
où σ_x et σ_y désignent respectivement les écarts-types de x et y.

Le coefficient de corrélation linéaire est compris entre -1 et 1 :

$$-1 \leq r(x,y) \leq 1 \tag{2.55}$$

Le coefficient de corrélation linéaire est :

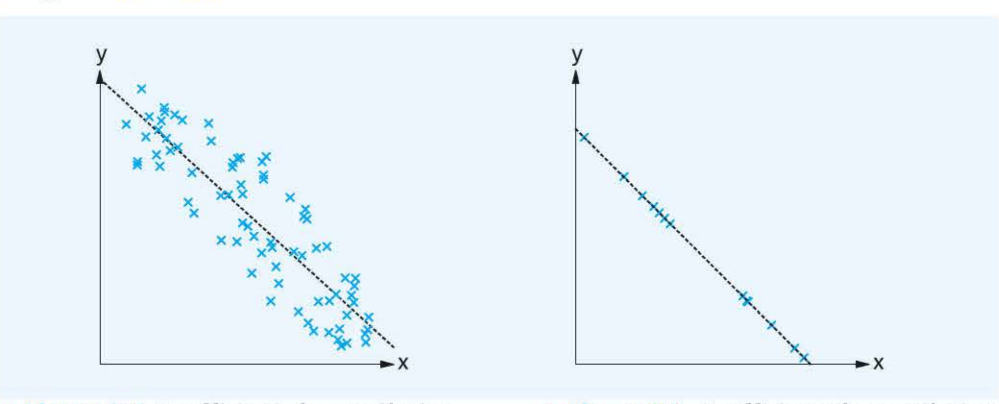
– positif si les variables x et y évoluent dans le même sens ; la corrélation étant d'autant plus forte que $r(x,y)$ est proche de 1. Les couples de points (x,y) sont alors concentrés autour d'une droite croissante, ainsi que cela est reproduit sur les figures 2.5 et 2.6.



▲ **Figure 2.5 Coefficient de corrélation linéaire positif**

▲ **Figure 2.6 Coefficient de corrélation linéaire égal à 1**

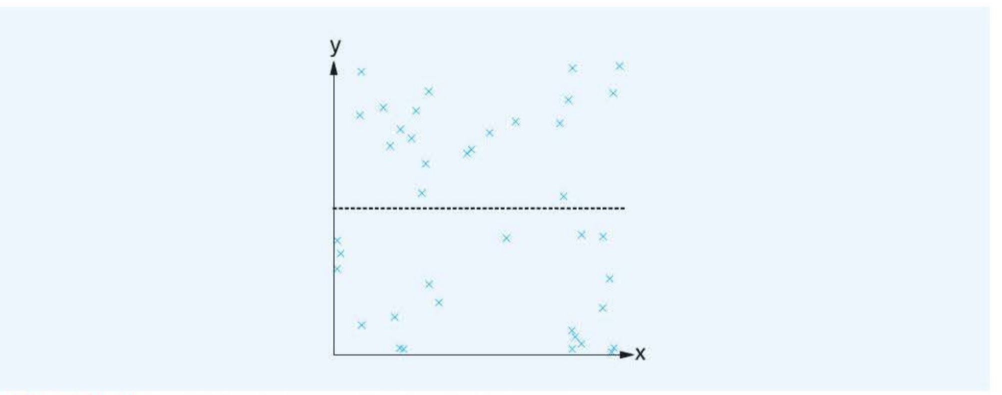
– négatif si les variables x et y évoluent en sens contraire ; la corrélation (négative) étant d'autant plus forte que $r(x,y)$ est proche de -1. Les couples de points (x,y) sont alors concentrés autour d'une droite décroissante, comme représenté sur les figures 2.7 et 2.8.



▲ **Figure 2.7 Coefficient de corrélation linéaire négatif**

▲ **Figure 2.8 Coefficient de corrélation linéaire égal à –1**

– nul si les variables x et y ne sont pas liées linéairement entre elles (▶ figure 2.9). Le nuage de points est alors dispersé dans le plan (x,y) et il n'est pas possible de l'ajuster par une droite autre que la droite horizontale $\hat{y} = b$ (parallèle à l'axe des abscisses).



▲ **Figure 2.9 Coefficient de corrélation linéaire nul**

Le calcul du coefficient de corrélation linéaire entre le taux d'investissement et le taux de croissance pour notre échantillon de 26 pays (▶ tableau 2.5) donne :

$$r(x,y) = \frac{12{,}14}{6{,}27 \times 2{,}16} = 0{,}90 \qquad (2.56)$$

$r(x,y)$ est ainsi relativement proche de 1, témoignant d'une corrélation positive importante entre le taux d'investissement et le taux de croissance économique dans notre échantillon.

Quelques précisions peuvent être apportées quant à l'interprétation du coefficient de corrélation linéaire :

- Un coefficient de corrélation linéaire égal à zéro n'implique pas nécessairement une absence de corrélation entre les deux variables étudiées. Il peut en effet exister une forte corrélation *non linéaire* entre les variables, qui ne peut être prise en compte, par définition, par une *droite* de régression.
- Une valeur élevée du coefficient de corrélation (proche de 1 en valeur absolue) n'implique pas nécessairement une forte dépendance entre les deux variables considérées. Il se peut en effet qu'une troisième variable agisse sur les deux variables étudiées, conduisant mécaniquement à une valeur élevée du coefficient de corrélation[5].
- Corrélation et **causalité** sont deux concepts différents qu'il convient de ne pas confondre : le calcul du coefficient de corrélation linéaire entre les variables x et y nous permet de dire si ces deux variables sont ou non liées entre elles, mais ne nous permet pas d'établir un lien de causalité[6]. En d'autres termes, il ne permet pas de déterminer si x cause (c'est-à-dire impacte) y ou si y cause x.

5 Un exemple illustratif peut être donné par les ventes de crèmes glacées et le nombre de noyades. Il serait incorrect de déduire, de l'observation d'une hausse des valeurs prises par ces deux variables durant l'été, l'existence d'une corrélation entre les deux séries. Ces deux variables ne sont en effet liées entre elles que par l'influence d'une troisième variable : la température, qui est plus élevée en été.

6 Cela peut en outre être illustré par le fait que deux droites de régression peuvent être associées au même coefficient de corrélation linéaire : la droite de régression de y sur x ($y = ax + b$) et la droite de régression de x sur y ($x = a'y + b'$).

3.2.2 Qualité d'une régression

Même si $r(x,y)$ nous permet d'apprécier le degré de corrélation linéaire entre x et y, il est possible de recourir à un indicateur plus général permettant de mesurer l'intensité de la liaison entre deux variables. Cet indicateur, appelé **coefficient de détermination**, reste valable en cas de corrélation non linéaire et permet en outre d'apprécier la qualité d'une régression. Son expression est issue de l'équation d'analyse de la variance.

Équation d'analyse de la variance. Ainsi que nous l'avons vu, la variance marginale est égale à la somme de la variance des moyennes conditionnelles et de la moyenne des variances conditionnelles. Considérons nos deux variables x et y pour un échantillon de i individus, $i = 1,...,N$. La variance marginale de y correspond à la **variance totale** (ou globale) de la série à une dimension, c'est-à-dire de la variable y. La variance des moyennes conditionnelles de y indique, par définition, la dispersion des moyennes conditionnelles entre elles. Il s'agit ainsi des moyennes des valeurs y_i pour chaque valeur de x. En d'autres termes, la variance des moyennes conditionnelles de y mesure la dispersion sur la droite de régression : c'est la **variance expliquée** par la régression. La moyenne des variances conditionnelles de y représente la dispersion moyenne des distributions conditionnelles de y : il s'agit de la dispersion moyenne des points autour de la droite de régression. En d'autres termes, c'est la variance qui n'est pas expliquée par la régression, on l'appelle **variance résiduelle**. On en déduit l'équation suivante :

$$\text{Variance totale} = \text{Variance expliquée} + \text{Variance résiduelle} \tag{2.57}$$

soit encore :

$$V(y) = V(\hat{y}) + V(e) \tag{2.58}$$

où $\hat{y}$ est la valeur située sur la droite de régression, $\hat{y} = ax + b$ (où a et b sont supposés estimés par la méthode des MCO), et e désigne l'écart entre la valeur observée de y et $\hat{y}$, $e = y - \hat{y}$. L'écart e porte également le nom de **résidu** et $\hat{y}$ de **valeur estimée** (ou ajustée) de y. L'équation (2.58) est connue sous le nom d'**équation d'analyse de la variance**.

La qualité d'une régression est ainsi d'autant meilleure que la variance expliquée par cette régression est élevée et, donc, que la variance résiduelle est faible. En d'autres termes, plus la variance expliquée est proche de la variance totale, meilleure est la régression. Il est possible de quantifier cela par le calcul d'un rapport, c'est-à-dire d'un nombre sans dimension, appelé coefficient de détermination.

Coefficient de détermination. Le coefficient de détermination se définit comme suit.

Définition 2.3

Le **coefficient de détermination**, noté R^2, est défini comme le rapport entre la variance expliquée et la variance totale (ou marginale) :

$$R^2 = \frac{\text{Variance expliquée}}{\text{Variance totale}} = 1 - \frac{\text{Variance résiduelle}}{\text{Variance totale}} \tag{2.59}$$

soit encore :

$$R^2 = \frac{V(\hat{y})}{V(y)} = 1 - \frac{V(e)}{V(y)} \tag{2.60}$$

Le coefficient de détermination mesure ainsi la part ou le pourcentage de variance expliquée par la régression : il fournit une mesure du pouvoir explicatif de la régression. On a par construction :

$$0 \leq R^2 \leq 1 \tag{2.61}$$

Trois principaux cas peuvent alors se présenter :

- Un coefficient de détermination égal à 0 signifie que la variance expliquée est nulle : la régression n'explique pas le nuage de points, les variables x et y ne sont pas liées entre elles.
- Un coefficient de détermination égal à 1 correspond au cas où la variance expliquée est égale à la variance totale : la régression explique parfaitement le lien entre x et y et la droite d'ajustement passe par *tous* les points du nuage.
- Dans le cas général, le coefficient de détermination prend une valeur située entre ces deux extrêmes. Plus la valeur de R^2 est proche de 1, plus la variance expliquée est proche de la variance totale et meilleure est la qualité de la régression.

Dans le cas d'une régression *linéaire* entre *deux* variables x et y, il est également possible d'exprimer le coefficient de détermination par la relation suivante, aisément utilisable en pratique :

$$R^2 = \frac{Cov(x,y)^2}{V(x)V(y)} \tag{2.62}$$

et correspondant au carré du coefficient de corrélation linéaire entre x et y : $R^2 = r^2(x,y)$.

Démonstration

Afin de démontrer la relation (2.62), reprenons l'équation (2.60) :

$$R^2 = \frac{V(\hat{y})}{V(y)} \tag{2.63}$$

On sait que $\hat{y} = ax + b$, d'où $V(\hat{y}) = V(ax + b) = a^2V(x)$ (d'après la propriété de la variance selon laquelle $V(ax + b) = V(aX) = a^2V(x)$). Par ailleurs, nous avons vu que $a = \frac{Cov(x,y)}{V(x)}$. On en déduit donc :

$$V(\hat{y}) = \frac{Cov(x,y)^2}{V(x)} \tag{2.64}$$

D'où :

$$R^2 = \frac{Cov(x,y)^2}{V(x)V(y)} \tag{2.65}$$

■

En utilisant la relation (2.62), le calcul du coefficient de détermination associé à la régression du taux de croissance sur le taux d'investissement (► tableau 2.5) nous donne :

$$R^2 = \frac{12{,}14^2}{39{,}11 \times 4{,}65} = 0{,}81 \tag{2.66}$$

On en déduit que le taux d'investissement explique 81 % de la variance du taux de croissance économique.

Les points **clés**

- Un tableau à double entrée permet de définir les distributions marginales et conditionnelles des deux variables étudiées, les distributions conditionnelles tenant compte simultanément de l'information contenue dans les deux variables.

- L'analyse de régression permet de quantifier le lien entre deux variables x et y en ajustant le nuage de points formé des valeurs du couple (x,y) par une droite, appelée droite de régression.

- Les coefficients de la droite de régression sont obtenus par la méthode des moindres carrés ordinaires.

- La coefficient de corrélation linéaire est un indicateur du degré de liaison linéaire entre deux variables.

- L'équation d'analyse de la variance et le coefficient de détermination sont utilisés pour juger de la qualité d'une régression.

ÉVALUATION

► Corrigés sur www.dunod.com

QCM

Pour chacune des questions suivantes, indiquer si les affirmations sont vraies ou fausses (il peut y avoir plusieurs réponses vraies pour chaque question).

1 Distributions marginales et conditionnelles

a. La distribution marginale de la variable x tient compte de l'information contenue dans la deuxième variable étudiée, y.

b. Les distributions conditionnelles de chacune des deux variables, x et y, croisent l'information contenue dans ces deux variables.

c. Contrairement aux fréquences marginales, la somme des fréquences conditionnelles est toujours égale à l'unité.

d. Les fréquences marginales et conditionnelles sont liées entre elles uniquement si les variables x et y sont dépendantes l'une de l'autre.

e. Dans un tableau à double entrée et sans prendre en compte la marge verticale, la somme colonne par colonne de l'ensemble des colonnes est égale à l'effectif total.

2 Liaison entre deux variables

a. Deux variables indépendantes sont telles que si l'une augmente, l'autre diminue.

b. Deux variables corrélées évoluent dans le même sens.

c. Les moyennes marginales et conditionnelles sont identiques pour chacune des deux variables étudiées si celles-ci sont indépendantes l'une de l'autre.

d. Les moyennes conditionnelles sont égales aux valeurs des variables dans le cas d'une liaison fonctionnelle entre x et y.

e. Lorsque y croît quand x décroît, on dit que les variables sont corrélées négativement.

3 Droite de régression et ajustement

a. La droite de régression passe par tous les couples (x,y) du nuage de points.

b. La droite de régression passe par le point de coordonnées $(\bar{x}, \bar{y})$.

c. La droite qui ajuste le mieux le nuage de points est telle que la somme des écarts entre les valeurs observées y_i de y et les valeurs $\hat{y}_i$ situées sur la droite est nulle, $i = 1,\ldots,N$.

d. Dans l'expression de la droite de régression, $\hat{y} = ax + b$, b désigne le coefficient de pente.

e. Les paramètres a et b de la droite de régression sont obtenus par la méthode des moindres carrés ordinaires.

4 Coefficient de corrélation linéaire

a. Un coefficient de corrélation linéaire égal à -1 témoigne d'une absence de corrélation entre les variables étudiées x et y.

b. Un coefficient de corrélation linéaire supérieur à 1 témoigne d'une très forte corrélation entre x et y.

c. Si le coefficient de corrélation linéaire $r(x,y)$ est égal à 1, on peut en déduire que x influence y.

d. Si $r(x,y) = 0$, les variables x et y n'ont aucun lien entre elles.

e. Plus un nuage de points a une forme allongée, plus la corrélation entre les deux variables est forte.

5 Analyse de la variance et coefficient de détermination

a. L'équation d'analyse de la variance est telle que la variance expliquée est égale à la différence entre la variance marginale et la variance résiduelle.

b. Le coefficient de détermination permet de juger de la qualité d'une régression.

c. Le coefficient de détermination est égal au rapport entre la variance résiduelle et la variance totale.

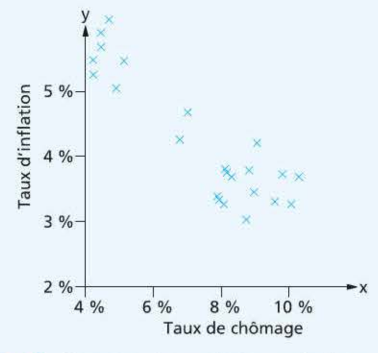
d. Le coefficient de détermination est compris entre -1 et 1.

e. Plus la variance résiduelle est faible, meilleure est la qualité d'une régression.

Exercice

6 Étude de la liaison taux de chômage/taux d'inflation

On considère un échantillon de 30 pays. On cherche à étudier, pour ces 30 pays, le lien entre le taux d'inflation et le taux de chômage. La figure 2.10 reproduit le nuage de points formé des couples (x,y) où x désigne le taux de chômage et y le taux d'inflation. On donne par ailleurs les statistiques suivantes, i désignant le pays $(i = 1,\ldots,30)$: $\bar{x} = 6{,}71$ et $\bar{y} = 4{,}74$; $V(x) = 4{,}64$ et $V(y) = 1{,}51$ et $\sum_{i=1}^{30} x_i y_i = 881{,}01$.



▲ **Figure 2.10 Nuage de points, relation taux de chômage/taux d'inflation**

1. Que peut-on dire du lien éventuel entre les deux variables au regard de la figure 2.10 ? Ce résultat est-il attendu d'un point de vue économique ?
2. Calculer la covariance entre les deux variables.
3. En déduire la valeur du coefficient de corrélation linéaire entre les deux variables.
4. Déterminer les valeurs des coefficients a et b de la droite de régression $\hat{y} = ax + b$.
5. Calculer la valeur du coefficient de détermination. Commenter.

Sujet d'examen

7 Université Paris Ouest, extrait

On considère trois séries mensuelles du taux de change du dollar (exprimées en logarithme) sur la période allant de janvier 1990 à avril 2004. On désigne par :

- EUR_t, le taux de change (au mois t) du dollar vis-à-vis de l'euro ;
- DKK_t, le taux de change (au mois t) du dollar vis-à-vis de la couronne danoise ;
- ATS_t, le taux de change (au mois t) du dollar vis-à-vis du schilling autrichien.

avec $t = 1,\ldots,172$. On donne par ailleurs diverses valeurs figurant dans le tableau 2.6.

▼ **Tableau 2.6 Statistiques sur les séries de taux de change**

Variable	Moyenne	Écart-type
EUR_t	−0,1227	0,1278
DKK_t	1,8942	0,1242
ATS_t	2,4998	0,1304

Somme des carrés	Somme des produits
$\sum_{t=1}^{172} EUR_t^2 = 5{,}3820$	$\sum_{t=1}^{172} EUR_t \times DKK_t = -37{,}3106$
$\sum_{t=1}^{172} DKK_t^2 = 619{,}7721$	$\sum_{t=1}^{172} EUR_t \times ATS_t = -49{,}9646$
$\sum_{t=1}^{172} ATS_t^2 = 1077{,}7743$	$\sum_{t=1}^{172} DKK_t \times ATS_t = 817{,}2116$

1. Calculer les coefficients de variation des trois séries. Commenter.
2. Calculer les coefficients de corrélation entre DKK et EUR d'une part, puis entre DKK et ATS d'autre part. Commenter les résultats obtenus.
3. On considère le modèle mettant en relation DKK et EUR.

   a. Déterminer, par la méthode des moindres carrés ordinaires, les valeurs des coefficients a et b de la droite de régression $\widehat{DKK}_t = aEUR_t + b$.

   b. Selon ce modèle, quel est l'effet d'une augmentation de 1 % du taux de change du dollar vis-à-vis de l'euro sur celui du dollar vis-à-vis de la couronne danoise ?

   c. On désigne par e_t le $t^{\text{ième}}$ résidu de l'ajustement obtenu et on donne : $\sum_{t=1}^{172} e_t^2 = 0{,}1169$. En écrivant l'équation d'analyse de la variance, déterminer la valeur de la variance expliquée. Quelle est la part de la variance résiduelle dans la variance totale ?

d. Calculer et interpréter la valeur du coefficient de détermination de la régression.

e. Que devient le coefficient a si l'on multiplie par 10 toutes les valeurs observées des variables EUR et DKK ? Même question si, au lieu de multiplier par 10, on ajoute 10 à chacune des valeurs observées des deux variables.

4. On s'intéresse désormais à la relation entre DKK et ATS. L'application de la méthode des MCO a conduit aux résultats suivants : $\widehat{DKK}_t = 0{,}9474 \times ATS_t - 0{,}4741$. Sachant que la variance de la variable ajustée de ce modèle est égale à 0,0153, calculer la valeur de la somme des carrés des résidus. Calculer le coefficient de détermination de cette nouvelle régression.

5. Comparer les résultats numériques obtenus pour les deux modèles qui font l'objet des questions 3 et 4 ci-dessus et conclure.

POUR ALLER PLUS LOIN

Terme d'erreur d'un modèle de régression

Nous avons considéré dans ce chapitre un modèle de régression linéaire faisant intervenir deux variables. Si l'on reprend le cas de la fonction de consommation keynésienne, on peut représenter celle-ci pour une date t donnée sous la forme du modèle $C_t = cR_t + c_0$ où C_t désigne la consommation (variable expliquée) à la date t, R_t est le revenu (variable explicative) à la même date, c et c_0 étant des paramètres. Cette écriture revient à supposer que la consommation est expliquée uniquement par le revenu. Si une telle relation est exacte, il est très aisé d'obtenir les valeurs des paramètres c et c_0 : il suffit en effet de disposer de deux observations et de les joindre au moyen d'une droite, les autres observations se situant sur cette même droite. Un tel schéma n'est cependant pas représentatif de la réalité économique et le fait que seul le revenu intervienne comme variable explicative dans le modèle peut sembler très restrictif dans la mesure où il est fort probable que d'autres variables expliquent la consommation (comme le taux de chômage, le taux d'intérêt, etc.). En l'absence d'information supplémentaire, on rajoute alors un terme ε_t qui représente l'ensemble des autres variables explicatives qui ne figurent pas dans le modèle. Le modèle s'écrit :

$$C_t = cR_t + c_0 + \varepsilon_t \quad (2.67)$$

Le terme ε_t est une *variable aléatoire* appelée **erreur** ou **perturbation**. Il s'agit de l'*erreur de spécification* du modèle, dans la mesure où elle rassemble toutes les variables, autres que le revenu, qui n'ont pas été prises en compte pour expliquer la consommation. Le terme d'erreur fournit ainsi une mesure de la différence entre les valeurs observées de la consommation et celles qui devraient être observées si le modèle était correctement spécifié. Notons que le terme d'erreur désigne non seulement l'erreur de spécification du modèle, mais peut également représenter une *erreur de mesure* liée à des problèmes de mesure des variables (explicatives) considérées.

Le terme d'erreur d'un modèle de régression n'est pas observable, mais il doit vérifier un certain nombre d'hypothèses afin que la méthode des moindres carrés ordinaires puisse être mise en oeuvre :

- Le terme d'erreur et la variable explicative ne sont pas liés entre eux. En d'autres termes, la variable explicative ne dépend pas du terme d'erreur, il s'agit d'une variable certaine au sens où elle est observée sans erreur.
- L'espérance mathématique de l'erreur est nulle. Cela revient à admettre, qu'en moyenne, le modèle est correctement spécifié et donc, qu'en moyenne, il n'y a pas d'erreur.
- La valeur du terme d'erreur à une date t ne dépend pas de sa valeur à une date t' avec $t \neq t'$. Cette hypothèse est celle de **non autocorrélation des erreurs**.
- La variance du terme d'erreur est constante sur l'ensemble de l'échantillon étudié. Cette hypothèse est connue sous le nom d'**homoscédasticité des erreurs**.

Chapitre 3

En économie et en finance, comme dans beaucoup d'autres domaines, les variables ou grandeurs représentatives des phénomènes étudiés varient au cours du temps et dans l'espace. Ainsi, les prix à la consommation fluctuent d'un mois sur l'autre (variation au cours du temps), mais également entre les pays (variation dans l'espace). Il en est de même du cours des actions, du pouvoir d'achat des ménages, du prix des logements, de la confiance des consommateurs, etc., qui sont des grandeurs qui varient au cours du temps et entre les pays, les régions, les catégories socio-professionnelles, etc.

Afin de comparer une même grandeur dans deux situations (dates, périodes, pays, régions, etc.), on choisit une de ces situations comme référence et l'on calcule un **indice**. Étant indépendant des unités relatives aux phénomènes étudiés, un indice permet de comparer l'évolution de grandeurs de natures différentes (prix des logements et salaire, par exemple).

LES GRANDS AUTEURS



Irving Fisher (1867-1947)

Irving Fisher est un économiste mathématicien américain, professeur à l'Université de Yale. Outre ses nombreuses contributions à la théorie économique sur les questions ayant trait à la valeur, le capital, les taux d'intérêt, la monnaie et l'instabilité financière, il est l'auteur en 1922 d'un ouvrage *The Making of Index Numbers* (Houghton Mifflin, Boston).

Dans ce livre, il présente l'**indice** qui porte son nom. Cet indice, défini comme une moyenne géométrique des indices de Laspeyres et de Paasche, a pour but de pallier les inconvénients de ces deux indices en représentant un compromis entre ceux-ci. ■

Indices

Plan

Pré-requis

- **Connaître** les opérations mathématiques de base et la notion de moyenne (▸ chapitre 1).
- **Savoir interpréter** un pourcentage.

Objectifs

- **Synthétiser** l'information contenue dans une grandeur économique en construisant un indice résumant l'évolution des prix et un indice résumant l'évolution des quantités relative à cette grandeur.
- **Comparer** l'évolution de grandeurs de natures différentes, ainsi que l'évolution d'une grandeur dans deux situations différentes.
- **Agréger** un ensemble de valeurs hétérogènes au sein d'un même indice.
- **Mettre en évidence** des variations de prix et de quantités liées à un changement de qualité.

Deux catégories d'indices peuvent être distinguées selon le type de grandeur étudiée. Ainsi, si l'on considère le prix d'un produit, la production d'une entreprise donnée, le cours de l'action d'une société particulière, il s'agit de **grandeurs simples** au sens où la grandeur est un nombre ne prenant qu'une seule valeur dans une situation donnée. Les indices calculés sur la base de ces grandeurs sont appelés **indices élémentaires**. En revanche, le niveau général des prix, la production industrielle, le cours des actions sont des **grandeurs complexes** dans la mesure où leur calcul nécessite d'agréger un ensemble de valeurs hétérogènes (prix des différents produits, production de diverses industries, cours de différentes actions). Les indices calculés sur la base de ces grandeurs sont appelés **indices synthétiques**.

1 Indices élémentaires

1.1 Quelques exemples introductifs

Considérons l'évolution du prix de vente annuel (en euros) du kilogramme de fraises en France de 1998 à 2013. Les valeurs, reportées dans le tableau 3.1, mettent en évidence une tendance haussière du prix de vente. Supposons que l'on souhaite quantifier cette hausse entre 2000 et 2013. À cette fin, il suffit d'effectuer le rapport suivant :

$$I_{2013/2000} = \frac{9,69}{5,61} = 1,7273 \qquad (3.1)$$

▼ **Tableau 3.1** **Prix de vente du kilogramme de fraises et SMIC brut**

Date	Prix de vente fraises (euros)	SMIC (euros par heure)
1998	5,29	
1999	5,36	
2000	5,61	
2001	6,55	6,53
2002	7,21	6,75
2003	7,92	7,01
2004	7,4	7,4
2005	7,84	7,82
2006	7,93	8,15
2007	7,93	8,36
2008	8,53	8,61
2009	8,28	8,77
2010	9,32	8,86
2011	9,51	9,02
2012	9,84	9,31
2013	9,69	9,43

Source : INSEE.

On en déduit que le prix du kilogramme de fraises a augmenté de 72,73 % entre 2000 et 2013 en France. Nous venons ici de comparer une même grandeur (prix de vente du kilogramme de fraises) à deux dates différentes.

Les données figurant dans le tableau 3.1 nous permettent également de comparer l'évolution du prix de vente des fraises à celle du SMIC horaire. Considérons, à titre d'exemple, l'évolution entre 2001 et 2013 de ces deux grandeurs et calculons les rapports suivants :

$$I^{fraises}_{2013/2001} = \frac{9{,}69}{6{,}55} = 1{,}4794 \qquad (3.2)$$

et :

$$I^{SMIC}_{2013/2001} = \frac{9{,}43}{6{,}53} = 1{,}4441 \qquad (3.3)$$

Il apparaît que le prix du kilogramme de fraises a augmenté légèrement plus fortement que le SMIC horaire entre 2001 et 2013, avec une hausse de 47,94 % pour le prix des fraises et de 44,41 % pour le SMIC. Ces calculs de rapports rendent ainsi aisée la comparaison de l'évolution de deux grandeurs différentes.

Comme nous l'avons précédemment mentionné, il est également possible de comparer une même grandeur dans deux catégories (pays, régions, entreprises, etc.) différentes à une même date. Le prix de vente du maïs (en euros par hectare) en 2012 est égal à 20,47 en Lituanie et il est de 17,49 en Pologne. Le rapport suivant :

$$I_{Lituanie/Pologne} = \frac{20{,}47}{17{,}49} = 1{,}1704 \qquad (3.4)$$

nous indique que le maïs coûte 17,04 % plus cher en Lituanie qu'en Pologne.

1.2 Définition

Les rapports calculés dans les exemples ci-dessus constituent des **indices élémentaires**[1].

Définition 3.1

Considérons une grandeur simple g prenant la valeur g_0 à la date 0 et la valeur g_t à la date t. On appelle **indice élémentaire** $I^g_{t/0}$ le nombre sans dimension suivant :

$$I^g_{t/0} = \frac{g_t}{g_0} \qquad (3.5)$$

La date 0 correspond à la **date de référence** ou **base** de l'indice, la date t étant la **date courante**.

1 Plus précisément, les indices relatifs au prix de vente de la fraise et au SMIC correspondent à des **indices élémentaires de variation** (ou d'évolution). Ces indices permettent de calculer l'évolution d'une même grandeur (prix de vente de la fraise, SMIC) entre deux dates différentes (2001 et 2013). Dans les exemples du prix de vente du maïs en 2012 en Lituanie et en Pologne, l'indice calculé est un **indice élémentaire de répartition**. Cet indice permet de réaliser des comparaisons entre des catégories différentes (Lituanie et Pologne) pour une même grandeur (prix de vente du maïs) et à une même date (2012). Dans la suite, nous parlerons simplement d'indice élémentaire.

Cet indice (de variation) mesure ainsi l'évolution de la grandeur g entre la date de base (date 0) et la date courante (date t)[2].

On exprime fréquemment un indice en pourcentage :

$$i^g_{t/0} = \frac{g_t}{g_0} \times 100 \tag{3.6}$$

L'indice à la date t défini dans l'équation (3.6) est ainsi dit exprimé base 100 à la date de référence. Notons que si l'on ne multiplie plus par 100 (équation (3.5)), on parle d'indice base 1 à la date 0.

1.3 Propriétés

Reprenons l'exemple précédent du prix du maïs en 2012 en Lituanie et en Pologne et supposons que ce prix soit, en 2012, 18 % plus cher en Grèce qu'en Lituanie. On peut donc écrire :

$$I_{Grèce/Lituanie} = 1{,}18 \tag{3.7}$$

Sachant que $I_{Lituanie/Pologne} = 1{,}1704$, il est possible de comparer le prix du maïs en Grèce et en Pologne :

$$I_{Grèce/Pologne} = I_{Grèce/Lituanie} \times I_{Lituanie/Pologne} = 1{,}18 \times 1{,}1704 = 1{,}3811 \tag{3.8}$$

On en déduit ainsi que le prix du maïs est 38,11 % plus cher en Grèce qu'en Pologne en 2012.

Cet exemple illustre la **propriété de circularité** ou de **transitivité** d'un indice élémentaire :

$$I^g_{t/0} = I^g_{t/t'} \times I^g_{t'/0} = \frac{g_t}{g'_t} \times \frac{g'_t}{g_0} = \frac{g_t}{g_0} \tag{3.9}$$

soit encore :

$$I^g_{t/t'} = \frac{I^g_{t/0}}{I^g_{t'/0}} \tag{3.10}$$

Cette propriété permet ainsi de comparer les dates 0 et t, d'une part, et les dates 0 et t', d'autre part, mais aussi les dates t et t' (équation (3.10)). Dans ce dernier cas, la date de référence est la date t' qui se substitue à la date 0, témoignant d'un changement de base. La propriété de circularité peut se généraliser comme suit :

$$I^g_{t/0} = I^g_{t/t-1} \times I^g_{t-1/t-2} \times \ldots \times I^g_{1/0} \tag{3.11}$$

et l'on parle alors d'**enchaînement** : on a une chaîne d'indices élémentaires, ces derniers étant dits enchaînables. Les indices élémentaires vérifient également la **propriété de réversibilité** selon laquelle :

$$I^g_{t/0} = \frac{1}{I^g_{0/t}} \tag{3.12}$$

2 Si l'on considère la grandeur simple g non plus aux dates 0 et t, mais pour deux catégories différentes (x et y) à la même date (par exemple à la date t), la relation $I^g_{y/x} = \frac{g_y}{g_x}$ définit l'indice élémentaire de répartition. La valeur de référence, ou base de l'indice, est ici la valeur g_x, c'est-à-dire la valeur de la grandeur g prise par la catégorie x à la date t. Cet indice de répartition mesure donc la variation relative de la grandeur g entre la valeur de référence (catégorie x) et la valeur de la catégorie y à la date t. Afin de ne pas alourdir les développements, on utilisera dans la suite du chapitre la notation générique $I^g_{t/0}$ où t désigne la valeur de référence de la grandeur g (date de référence pour l'indice de variation, catégorie de référence pour l'indice de répartition).

Ainsi, sachant que le prix du maïs est 38,11 % plus cher en Grèce qu'en Pologne ($I_{Grèce/Pologne} = 1{,}3811$), on en déduit que :

$$I_{Pologne/Grèce} = \frac{1}{I_{Grèce/Pologne}} = \frac{1}{1{,}3811} = 0{,}7241 \qquad (3.13)$$

c'est-à-dire que le prix du maïs est 27,59 % moins élevé en Pologne qu'en Grèce.

Remarque : Un indice élémentaire vérifie également les propriétés suivantes :

– Supposons que la grandeur g soit telle que $g = a \times b$ où a et b sont deux grandeurs. On a alors :

$$I^g_{t/0} = I^{ab}_{t/0} = I^a_{t/0} \times I^b_{t/0} \qquad (3.14)$$

– Supposons que la grandeur g soit telle que $g = \frac{a}{b}$ où a et b sont deux grandeurs. On a alors :

$$I^g_{t/0} = I^{a/b}_{t/0} = \frac{I^a_{t/0}}{I^b_{t/0}} \qquad (3.15)$$

2 Indices synthétiques

En économie, les grandeurs étudiées sont souvent complexes, c'est-à-dire composées de plusieurs grandeurs simples qu'il faut agréger ou synthétiser. Ainsi, l'indice des prix à la production dans l'industrie fourni par l'INSEE est calculé sur la base des prix de 24 000 produits. Il s'agit d'un **indice synthétique** résumant – c'est-à-dire agrégeant – les 24 000 indices élémentaires relatifs au prix de chacun des produits considérés. Dans la mesure où il existe différentes façons d'agréger une série d'indices élémentaires, il existe plusieurs indices synthétiques.

2.1 Indices de valeur, des prix et de volume : généralités

En économie, on étudie souvent l'évolution des prix, des quantités et de leur produit, appelé valeur. Trois types d'indices peuvent alors être calculés : **indice des prix**, **indice des quantités** (ou **indice de volume**) et **indice de valeur**. Considérons une grandeur complexe g composée de k éléments : $g^1, g^2, \ldots, g^k$. Ainsi, si g est l'indice des prix à la production, les éléments g^i, $i = 1, \ldots, k$, désignent les 24 000 produits utilisés dans le calcul de l'indice. Soient par ailleurs les notations suivantes : p^i_0 et p^i_t le prix respectivement aux dates 0 et t d'un élément g^i, et q^i_0 et q^i_t les quantités respectivement aux dates 0 et t de ce même élément g^i.

Définition 3.2

L'**indice de valeur** $I^V_{t/0}$ est donné par le rapport entre la somme des valeurs des k éléments g^i, $i = 1, \ldots, k$, de la grandeur g considérée à la date t et cette même somme à la date de référence 0, soit[3] :

$$I^V_{t/0} = \frac{\Sigma_i p^i_t q^i_t}{\Sigma_i p^i_0 q^i_0} \qquad (3.16)$$

3 Afin d'alléger les notations, on utilisera la convention suivante dans tout ce chapitre : $\sum_i = \sum_{i=1}^{k}$.

FOCUS

L'indice de valeur

Notons g^i un élément rentrant dans la composition de la grandeur complexe g, V_t^i sa valeur (monétaire) à la date t, p_t^i son prix unitaire à la date t et q_t^i sa quantité à la date t. À chaque date, la valeur de l'élément g^i est égale au produit de son prix unitaire et de la quantité correspondante, soit $V_t^i = p_t^i q_t^i$ pour la date t et $V_0^i = p_0^i q_0^i$ pour la date 0. À chaque date, la valeur de la grandeur complexe g est égale à la somme des valeurs des éléments g^i la composant, soit $V_t = \Sigma_i V_{it} = \Sigma_i p_t^i q_t^i$ pour la date t et $V_0 = \Sigma_i V_{i0} = \Sigma_i p_0^i q_0^i$ pour la date 0. L'indice de valeur (monétaire) de la grandeur complexe g entre la date t et la date 0, $I_{t/0}^V$, est donc égal à (équation (3.16)) :

$$I_{t/0}^V = I_{t/0}^{pq} = \frac{\Sigma_i V_t^i}{\Sigma_i V_0^i} = \frac{\Sigma_i p_t^i q_t^i}{\Sigma_i p_0^i q_0^i} \tag{3.17}$$

Cet indice de valeur est relativement peu informatif au sens où, s'il augmente, il n'est pas possible de distinguer si cette hausse provient d'une augmentation des prix accompagnée d'une baisse des quantités ou de toute autre combinaison. Pour pallier cette difficulté, on suppose que l'une des deux variables (prix ou quantité) est fixe alors que l'autre varie en calculant des indices de prix et de quantités :

– On calcule un indice des prix en neutralisant l'influence des quantités, c'est-à-dire que l'on considère que les quantités sont fixes sur la période considérée.

– De façon réciproque, un indice des quantités (ou indice de volume) se calcule en neutralisant l'influence des prix, c'est-à-dire que l'on considère que les prix sont constants sur la période considérée.

Les indices des prix et des quantités les plus fréquemment utilisés sont les indices de Laspeyres et de Paasche que nous présentons ci-après.

2.2 Indices de Laspeyres et de Paasche

2.2.1 Définitions générales

Notons α^i le poids de l'élément g^i ($i = 1,\ldots,k$) dans la grandeur complexe g. Les coefficients α^i ($i = 1,\ldots,k$) sont donc des **coefficients de pondération**. Ainsi, si l'on considère l'indice des prix à la consommation des ménages, ces coefficients représentent la part de chaque bien et service dans la consommation des ménages ; il peut s'agir par exemple de la part des dépenses de loyer, la part des dépenses de consommation de poisson, etc. En d'autres termes, dans le cas d'indices des prix, il s'agit de coefficients budgétaires. Par définition, on a donc :

$$\Sigma_i \alpha_0^i = \Sigma_i \alpha_t^i = 1 \tag{3.18}$$

Les indices de Laspeyres et de Paasche sont des moyennes pondérées par ces coefficients α^i des indices élémentaires $I_{t/0}^i$ relatifs à chaque élément g^i, $i = 1,\ldots,k$, de la grandeur g.

Définition 3.3

L'**indice de Laspeyres** $L^g_{t/0}$ est la moyenne arithmétique des indices élémentaires pondérée par les coefficients de la date de référence (α^i_0) :

$$L^g_{t/0} = \Sigma_i \alpha^i_0 I^i_{t/0} = \Sigma_i \alpha^i_0 \frac{g^i_t}{g^i_0} \tag{3.19}$$

Définition 3.4

L'**indice de Paasche** $P^g_{t/0}$ est la moyenne harmonique des indices élémentaires pondérée par les coefficients de la date courante (α^i_t) :

$$P^g_{t/0} = \frac{1}{\Sigma_i \frac{\alpha^i_t}{I^i_{t/0}}} = \frac{1}{\Sigma_i \alpha^i_t \frac{g^i_0}{g^i_t}} \tag{3.20}$$

Remarque : Il est important de souligner que les indices élémentaires entrant dans le calcul de l'indice synthétique doivent être basés (1 ou 100) à la même date.

2.2.2 Indice de Laspeyres

Indice des prix de Laspeyres. Dans le cas de l'indice de Laspeyres, les coefficients de pondération sont ceux de la date de référence et sont donnés par :

$$\alpha^i_0 = \frac{p^i_0 q^i_0}{\Sigma_i p^i_0 q^i_0} \tag{3.21}$$

D'après l'équation (3.19), on peut écrire l'indice des prix de Laspeyres $L^g_{t/0}(p)$ comme suit :

$$L^g_{t/0}(p) = \Sigma_i \alpha^i_0 \frac{p^i_t}{p^i_0} = \Sigma_i \frac{p^i_0 q^i_0}{\Sigma_i p^i_0 q^i_0} \frac{p^i_t}{p^i_0} \tag{3.22}$$

D'où :

$$L^g_{t/0}(p) = \frac{\Sigma_i p^i_t q^i_0}{\Sigma_i p^i_0 q^i_0} \tag{3.23}$$

Ainsi, les quantités sont constantes (ce sont celles de la date de référence puisque nous sommes dans le cas d'un indice de Laspeyres) et seuls les prix varient puisqu'il s'agit d'un indice des prix. Cet indice s'interprète comme le rapport entre la dépense totale à la date de référence évaluée aux prix courants et la dépense totale de la date de référence.

Si l'on prend l'exemple de l'indice des prix à la consommation, l'indice des prix de Laspeyres décrit ainsi l'évolution du prix d'un panier de biens dont la composition est restée fixe entre les deux dates et est celle de la date de référence.

Indice de volume de Laspeyres. Les coefficients de pondération étant donnés par l'équation (3.21), on peut écrire en utilisant la relation (3.19) l'indice de volume de Laspeyres :

$$L^g_{t/0}(q) = \Sigma_i \alpha^i_0 \frac{q^i_t}{q^i_0} = \Sigma_i \frac{p^i_0 q^i_0}{\Sigma_i p^i_0 q^i_0} \frac{q^i_t}{q^i_0} \tag{3.24}$$

$$L^g_{t/0}(q) = \frac{\Sigma_i p^i_0 q^i_t}{\Sigma_i p^i_0 q^i_0} \tag{3.25}$$

Les prix sont constants (ce sont ceux de la date de référence puisque nous sommes dans le cas d'un indice de Laspeyres) et seules les quantités varient puisqu'il s'agit d'un indice de volume. Cet indice s'interprète comme le rapport entre la dépense totale à la date courante évaluée aux prix à la date de référence et la dépense totale de la date de référence.

2.2.3 Indice de Paasche

Indice des prix de Paasche. Dans le cas de l'indice de Paasche, les coefficients de pondération sont ceux de la date courante et sont donnés par :

$$\alpha^i_t = \frac{p^i_t q^i_t}{\Sigma_i p^i_t q^i_t} \tag{3.26}$$

D'après l'équation (3.20), on peut écrire l'indice des prix de Paasche $P^g_{t/0}(p)$ comme suit :

$$P^g_{t/0}(p) = \frac{1}{\Sigma_i \alpha^i_t \frac{p^i_0}{p^i_t}} = \frac{1}{\Sigma_i \frac{p^i_t q^i_t}{\Sigma_i p^i_t q^i_t} \frac{p^i_0}{p^i_t}} \tag{3.27}$$

D'où :

$$P^g_{t/0}(p) = \frac{\Sigma_i p^i_t q^i_t}{\Sigma_i p^i_0 q^i_t} \tag{3.28}$$

Ainsi, les quantités sont constantes (ce sont celles de la date courante puisque nous sommes dans le cas d'un indice de Paasche) et seuls les prix varient puisqu'il s'agit d'un indice des prix. Cet indice s'interprète comme le rapport entre la dépense totale à la date courante et la dépense totale de la date courante évaluée aux prix de la date de référence. Si l'on reprend l'exemple de l'indice des prix à la consommation, l'indice des prix de Paasche décrit ainsi l'évolution du prix d'un panier de biens dont la composition est celle de la date courante.

Indice de volume de Paasche. Les coefficients de pondération étant donnés par l'équation (3.26), on peut écrire en utilisant la relation (3.20) l'indice de volume de Paasche :

$$P^g_{t/0}(q) = \frac{1}{\Sigma_i \alpha^i_t \frac{q^i_0}{q^i_t}} = \frac{1}{\Sigma_i \frac{p^i_t q^i_t}{\Sigma_i p^i_t q^i_t} \frac{q^i_0}{q^i_t}} \tag{3.29}$$

$$P^g_{t/0}(q) = \frac{\Sigma_i p^i_t q^i_t}{\Sigma_i p^i_t q^i_0} \tag{3.30}$$

Les prix sont constants (ce sont ceux de la date courante puisque nous sommes dans le cas d'un indice de Paasche) et seules les quantités varient puisqu'il s'agit d'un indice de volume. Cet indice s'interprète comme le rapport entre les quantités à la date courante évaluées aux prix à cette même date et la dépense totale de la date de référence évaluée aux prix courants.

2.3 Indice de Fisher

Dans la mesure où il n'existe pas de critère permettant de conclure à la supériorité d'un des deux indices précédents – Laspeyres et Paasche – par rapport à l'autre, l'idée consiste à construire un indice, appelé **indice de Fisher**, représentant une combinaison des indices de Laspeyres et Paasche dont la valeur se situe « entre » celles de ces deux indices.

Définition 3.5

L'**indice de Fisher** est défini comme la moyenne géométrique des indices de Laspeyres et Paasche, soit :

$$F^g_{t/0} = \sqrt{L^g_{t/0} P^g_{t/0}} \quad (3.31)$$

Il est alors possible de définir :

- Un indice des prix de Fisher :

$$F^g_{t/0}(p) = \sqrt{L^g_{t/0}(p) P^g_{t/0}(p)} \quad (3.32)$$

- Un indice de volume de Fisher :

$$F^g_{t/0}(q) = \sqrt{L^g_{t/0}(q) P^g_{t/0}(q)} \quad (3.33)$$

FOCUS

Quel indice synthétique privilégier ?

Ainsi que nous l'avons précédemment souligné, il n'existe pas de critère général permettant de statuer sur la supériorité d'un indice synthétique par rapport à un autre. Il est cependant possible de présenter les principaux avantages et inconvénients de ceux-ci. Supposons que l'on étudie l'évolution de la consommation d'un panier composé de plusieurs biens. Dans le cas de l'indice de Laspeyres, les coefficients de pondération sont fixes, c'est-à-dire que l'on suppose que la structure de la consommation ne se modifie pas sur la période étudiée.

En conséquence, si l'on considère que les coefficients de pondération sont fixés à la date de référence, plus la date courante est éloignée de la date initiale, plus il est probable que la structure du panier de biens du consommateur se soit modifiée et plus le risque que les coefficients de pondération soient obsolètes est important. Pour cette raison, le principal inconvénient attribué à l'indice de Laspeyres est qu'il tend à surestimer l'effet de l'évolution des prix sur le pouvoir d'achat du consommateur dans la mesure où il ne tient pas compte d'éventuelles substitutions entre les biens du panier considéré. Notons que cet inconvénient a pour conséquence que les coefficients de pondération de l'indice de Laspeyres sont révisés de façon périodique.

Dans le cas de l'indice de Paasche, les coefficients de pondération sont ceux de la date courante. Ceux-ci évoluent donc avec les prix, c'est-à-dire que la part des différents biens au sein du panier considéré évolue en même temps que les prix. Le calcul de l'indice de Paasche nécessite en conséquence de disposer simultanément des données relatives aux prix et aux quantités à chaque

date considérée (et non plus seulement des prix comme dans le cas de l'indice de Laspeyres). Le principal inconvénient tient ici en une difficulté de calcul supplémentaire liée à la disponibilité des données, expliquant pourquoi l'indice de Laspeyres est plus fréquemment utilisé que l'indice de Paasche[4]. Du fait de la variabilité des coefficients de pondération, l'indice de Paasche tend, au contraire de l'indice de Laspeyres, à sous-estimer l'effet de l'évolution des prix sur le pouvoir d'achat du consommateur. Il est important de souligner que les modifications de la structure de consommation ne dépendent évidemment pas que de l'évolution des prix relatifs des biens composant le panier. Il convient aussi de tenir compte de l'évolution des élasticités revenu.

Ainsi, même si le prix d'un bien diminue, cela n'implique évidemment pas que sa part dans la consommation va augmenter : pour certains biens dont le prix relatif a diminué, une hausse de revenu peut se traduire par une baisse de leur consommation relative (cas des biens inférieurs par exemple). Au total, il n'existe donc aucun critère théorique permettant de préférer un système de pondération par rapport à un autre, tout dépend de la façon dont les structures de consommation évoluent. Pour finir, notons que si les différences entre les pondérations des indices de Laspeyres et de Paasche sont faibles, on a la relation suivante :

$$P^g_{t/0} \leq F^g_{t/0} \leq L^g_{t/0} \qquad (3.34)$$

Cette inégalité s'obtient en notant que (i) sous l'hypothèse de coefficients de pondération égaux entre les deux indices, les indices de Laspeyres et de Paasche sont respectivement des moyennes arithmétique et harmonique d'indices élémentaires et (ii) la moyenne harmonique d'une série est toujours inférieure à la moyenne arithmétique de cette même série. Il est donc important de souligner qu'en cas de différences non négligeables entre les coefficients de pondération, cette inégalité n'a plus lieu d'être vérifiée.

2.4 Application empirique

Supposons que l'on souhaite calculer un indice synthétique du prix et des quantités d'un bouquet énergétique composé de pétrole, gaz naturel et charbon pour l'année 2012. L'année de référence retenue est 1990. On dispose, pour ces deux dates, du prix et de la quantité consommée en France pour chacune des trois énergies. Ces données sont reportées dans le tableau 3.2.

▼ Tableau 3.2 Pétrole, gaz naturel et charbon : prix et quantités consommées en France

Énergie i	Prix en 1990 p^i_0	Prix en 2012 p^i_t	Quantité en 1990 q^i_0	Quantité en 2012 q^i_t
Pétrole	24,5	94,13	1895	1687
Gaz naturel	1,64	2,76	29,3	42,5
Charbon	43,48	92,5	19,7	11,4

Source : British Petroleum (BP) Statistical Review of World Energy (juin 2013). Les prix sont exprimés en dollars par baril pour le pétrole, en dollars par millions d'unités thermales pour le gaz et en dollars par tonne pour le charbon. Les consommations sont exprimées en milliers de barils par jour pour le pétrole, en milliards de mètres cubiques pour le gaz et en millions de tonnes équivalent pétrole pour le charbon.

4 À titre d'exemple, les indices de prix à la consommation, de la production industrielle, de chiffre d'affaires, ou encore de prix à la production – pour n'en citer que quelques uns – calculés par l'INSEE sont des indices de Laspeyres.

À partir des données figurant dans le tableau 3.2, on peut calculer les indices élémentaires des prix $I^i_{t/0} = \frac{p^i_t}{p^i_0}$ en les exprimant base 1 en 1990. On constate ainsi au regard du tableau 3.3 qu'entre 1990 et 2012, les prix du pétrole ont augmenté de 284,2 %, ceux du gaz naturel de 68,3 % et ceux du charbon de 112,7 %. Le calcul des indices synthétiques nous donne les résultats suivants :

- Pour les indices de Laspeyres :

$$L^g_{t/0}(p) = \frac{\Sigma_i p^i_t q^i_0}{\Sigma_i p^i_0 q^i_0} = \frac{180279,47}{47332,11} = 3,8088 \quad (3.35)$$

et

$$L^g_{t/0}(q) = \frac{\Sigma_i p^i_0 q^i_t}{\Sigma_i p^i_0 q^i_0} = \frac{41896,87}{47332,11} = 0,8852 \quad (3.36)$$

- Pour les indices de Paasche :

$$P^g_{t/0}(p) = \frac{\Sigma_i p^i_t q^i_t}{\Sigma_i p^i_0 q^i_t} = \frac{159969,11}{41896,87} = 3,8182 \quad (3.37)$$

et

$$P^g_{t/0}(q) = \frac{\Sigma_i p^i_t q^i_t}{\Sigma_i p^i_t q^i_0} = \frac{159969,11}{180279,47} = 0,8873 \quad (3.38)$$

- Pour les indices de Fisher :

$$F^g_{t/0}(p) = \sqrt{L^g_{t/0}(p) P^g_{t/0}(p)} = \sqrt{3,8088 \times 3,8182} = 3,8135 \quad (3.39)$$

et

$$F^g_{t/0}(q) = \sqrt{L^g_{t/0}(q) P^g_{t/0}(q)} = \sqrt{0,8852 \times 0,8873} = 0,8863 \quad (3.40)$$

Ces calculs montrent ainsi que les prix du bouquet énergétique considéré ont augmenté de 280,9 % selon l'indice de Laspeyres et de 281,8 % selon l'indice de Paasche, la hausse se situant entre ces deux valeurs selon l'indice de Fisher avec une augmentation des prix de 281,4 %. S'agissant des volumes, on relève une baisse de la quantité consommée entre 1990 et 2012 de 11,5 % selon l'indice de Laspeyres, 11,3 % selon l'indice de Paasche et 11,4 % selon l'indice de Fisher. Les calculs des coefficients de pondération nous montrent en outre que la part des dépenses énergétiques consacrée à la consommation de pétrole est la plus élevée parmi les trois types d'énergie considérées, cette part s'élevant à 98,09 % en 1990 et 99,27 % en 2012.

▼ Tableau 3.3 Pétrole, gaz naturel et charbon : calcul des indices élémentaires et synthétiques

Énergie i	$I^i_{t/0}$	$p^i_0 q^i_0$	$p^i_t q^i_0$	α_0	$p^i_0 q^i_t$	$p^i_t q^i_t$	α_t
Pétrole	3,8420	46427,50	178376,35	0,9809	41331,50	158797,31	0,9927
Gaz naturel	1,6829	48,05	80,87	0,0010	69,70	117,30	0,0007
Charbon	2,1274	856,56	1822,25	0,0181	495,67	1054,50	0,0066
Somme		47332,11	180279,47	1	41896,87	159969,11	1

2.5 Propriétés des indices de Laspeyres, Paasche et Fisher

2.5.1 Circularité

Les indices de Laspeyres, Paasche et Fisher ne vérifient pas la propriété de circularité, ce qui constitue évidemment un inconvénient notable et implique qu'un changement de base nécessite de réeffectuer les différents calculs voulus :

$$L^g_{t/0} \neq L^g_{t/t'} \times L^g_{t'/0} \tag{3.41}$$

$$P^g_{t/0} \neq P^g_{t/t'} \times P^g_{t'/0} \tag{3.42}$$

$$F^g_{t/0} \neq F^g_{t/t'} \times F^g_{t'/0} \tag{3.43}$$

Démonstration

Effectuons, à titre d'exemple, la démonstration dans le cas de l'indice des prix de Laspeyres. On a :

$$\frac{L^g_{t/0}(p)}{L^g_{t'/0}(p)} = \frac{\Sigma_i p^i_t q^i_0}{\Sigma_i p^i_0 q^i_0} \times \frac{\Sigma_i p^i_0 q^i_0}{\Sigma_i p^i_{t'} q^i_0} = \frac{\Sigma_i p^i_t q^i_0}{\Sigma_i p^i_{t'} q^i_0} \tag{3.44}$$

et :

$$L^g_{t/t'}(p) = \frac{\Sigma_i p^i_t q^i_{t'}}{\Sigma_i p^i_{t'} q^i_{t'}} \tag{3.45}$$

On en déduit donc que $L^g_{t/t'}(p) \neq \dfrac{L^g_{t/0}(p)}{L^g_{t'/0}(p)}$, mettant en avant le non respect de la propriété de circularité. ■

2.5.2 Réversibilité

Les indices de Laspeyres et de Paasche ne sont pas réversibles :

$$L^g_{t/0} \neq \frac{1}{L^g_{0/t}} \tag{3.46}$$

$$P^g_{t/0} \neq \frac{1}{P^g_{0/t}} \tag{3.47}$$

En revanche, on a les relations suivantes entre ces deux indices :

$$L^g_{t/0} = \frac{1}{P^g_{0/t}} \tag{3.48}$$

$$P^g_{t/0} = \frac{1}{L^g_{0/t}} \tag{3.49}$$

L'indice de Fisher est quant à lui réversible :

$$F^g_{t/0} = \frac{1}{F^g_{0/t}} \tag{3.50}$$

2.5.3 Agrégation

Supposons que l'on étudie les dépenses des ménages et que l'on agrège ces dépenses par groupes : logement, consommation de viande, consommation de légumes,

consommation de produits laitiers, etc. On calcule, pour chacun de ces groupes, les indices de Laspeyres et de Paasche. L'indice global de l'ensemble des dépenses des ménages peut alors être obtenu à partir de ces indices calculés par groupes : l'indice global de Laspeyres (respectivement de Paasche) est en effet égal à la moyenne pondérée des indices de Laspeyres (respectivement de Paasche) calculés sur les différents groupes. Les indices de Laspeyres et de Paasche vérifient ainsi la propriété d'agrégation. L'indice de Fisher n'ayant pas une structure de moyenne arithmétique, il ne vérifie pas cette propriété.

2.5.4 Reconstitution de l'indice de valeur

Rappelons que l'indice de valeur est donné par :

$$I_{t/0}^{V} = I_{t/0}^{pq} = \frac{\Sigma_i p_t^i q_t^i}{\Sigma_i p_0^i q_0^i} \tag{3.51}$$

ce que l'on peut encore décomposer comme suit :

$$I_{t/0}^{V} = \frac{\Sigma_i p_t^i q_t^i}{\Sigma_i p_t^i q_0^i} \times \frac{\Sigma_i p_t^i q_0^i}{\Sigma_i p_0^i q_0^i} = \frac{\Sigma_i p_t^i q_t^i}{\Sigma_i p_0^i q_t^i} \times \frac{\Sigma_i p_0^i q_t^i}{\Sigma_i p_0^i q_0^i} \tag{3.52}$$

D'où les relations suivantes :

$$I_{t/0}^{V} = P_{t/0}^{g}(q) \times L_{t/0}^{g}(p) = P_{t/0}^{g}(p) \times L_{t/0}^{g}(q) \tag{3.53}$$

On en déduit que :

$$I_{t/0}^{V} = \sqrt{P_{t/0}^{g}(q) \times L_{t/0}^{g}(p) \times P_{t/0}^{g}(p) \times L_{t/0}^{g}(q)} \tag{3.54}$$

soit encore :

$$I_{t/0}^{V} = \sqrt{L_{t/0}^{g}(p) \times P_{t/0}^{g}(p)} \times \sqrt{L_{t/0}^{g}(q) \times P_{t/0}^{g}(q)} \tag{3.55}$$

d'où :

$$I_{t/0}^{V} = F_{t/0}^{g}(p) \times F_{t/0}^{g}(q) \tag{3.56}$$

L'indice de valeur est ainsi égal au produit des indices de prix et de volume de Fisher.

3 Raccords d'indices et indices chaînes

3.1 Raccords d'indices

Les grandeurs économiques telles que le PIB, la dépense de consommation des ménages, etc., sont calculées selon divers critères et nomenclatures, comme le Système des comptes nationaux ou le Système européen de comptes, régis au niveau international afin de rendre plus aisées les comparaisons entre pays. Or, ces systèmes et nomenclatures évoluent au cours du temps afin de s'adapter aux modifications de l'environnement et du fonctionnement économique. Des changements de base des indices peuvent alors être opérés afin de tenir compte de ces modifications en intégrant les

nouveaux éléments inclus régulièrement dans les nomenclatures. Il est ainsi fréquent lorsque l'on étudie l'évolution d'une grandeur économique sur une longue période d'avoir à gérer de tels changements de base dans les indices.

FOCUS

Choix de la période de base d'un indice

Le choix de la période de base ou de référence d'un indice (période 0) revêt une importance particulière dans la mesure où l'évolution de l'indice dépend de ce choix. En pratique, il convient d'éviter de retenir une période atypique, c'est-à-dire caractérisée par des fluctuations exceptionnelles (accidentelles ou saisonnières) donnant par exemple un poids inhabituel à un certain indice élémentaire, afin de ne pas fausser l'évolution de l'indice synthétique.

À cette fin, on s'efforce de choisir une date « normale » comme date de référence – c'est-à-dire dénuée d'événements exceptionnels – où l'on retient non pas une date particulière, mais une période de base composée de plusieurs dates. On peut ainsi choisir comme période de base une moyenne sur plusieurs années (pour un indice annuel) ou plusieurs mois (pour un indice mensuel) afin de lisser les effets d'éventuelles évolutions atypiques.

Un exemple caractéristique est celui des produits agricoles : les prix et quantités de ceux-ci étant fortement influencés par les conditions climatiques, il est d'usage de retenir comme base non pas une récolte correspondant à une date donnée, mais une moyenne sur plusieurs récoltes afin d'atténuer les effets d'événements climatiques ou conjoncturels exceptionnels.
Comme nous l'avons précédemment souligné, des changements de base sont en outre régulièrement opérés afin de prendre en compte les évolutions de l'environnement économique. Ainsi, plus on s'éloigne de la période de base, plus celle-ci devient obsolète à des fins de comparaison temporelle au sens où la structure du phénomène étudié (structure de la consommation, de la production, des échanges, etc.) s'est modifiée au cours du temps. Il convient alors d'actualiser la base (ce qui correspond à une actualisation des pondérations dans le cas d'un indice de Laspeyres), *via* un changement de base, en procédant à des raccords d'indices. Cette opération permet ainsi de tenir compte non seulement des modifications de la structure de la grandeur étudiée, mais également d'inclure de nouveaux produits, de supprimer des produits devenus obsolètes, de prendre en compte une nouvelle nomenclature, etc. Pour ces diverses raisons, les instituts de statistique tels que l'INSEE procèdent régulièrement à des changements de base des indices. À titre d'exemple, l'indice de la production industrielle calculé par l'INSEE subit un changement de base tous les cinq ans.

Considérons un indice I^g, base 1 à la date 0, calculé pour la grandeur g jusqu'à la date d, date à laquelle il est remplacé par un indice J^g base 1 à la date f[5]. Afin d'étudier l'évolution de la grandeur g sur l'ensemble de la période allant des dates 0 à t (avec $t > d$), il convient de procéder à un **raccord d'indices**, c'est-à-dire de déterminer la valeur qu'aurait pris l'indice I^g à la date t. Notons I'^g l'indice ainsi raccordé, on a :

$$I'^g_{t/0} = J^g_{t/f} \times \frac{I^g_{k/0}}{J^g_{k/f}} \quad (3.57)$$

5 Le raisonnement est strictement identique si l'on considère les indices en base 100 à la place de la base 1.

avec $f \leq k \leq d$. Le rapport $\frac{I^g_{k/0}}{J^g_{k/f}}$ est appelé **coefficient de raccordement** et correspond au coefficient par lequel on doit multiplier le nouvel indice afin d'en déduire la valeur prise par le précédent indice s'il avait continué à être calculé. Le choix de la date k est laissé au statisticien, mais l'on retient fréquemment la dernière date pour laquelle l'ancien indice est disponible.

Afin d'illustrer l'opération de raccord d'indices, considérons l'indice des prix de production des services de télécommunications français aux entreprises françaises. Cet indice, base 1 au premier trimestre 2007, est calculé par l'INSEE jusqu'au troisième trimestre 2012. À cette date, il s'élève à 0,833. À compter du quatrième trimestre 2012, il est remplacé par l'INSEE par un nouvel indice, base 1 en 2010 qui vaut 0,979 au troisième trimestre 2012 et s'élève à 0,918 au troisième trimestre de l'année 2013. Afin de calculer la valeur qu'aurait pris l'ancien indice au troisième trimestre de l'année 2013, on calcule le coefficient de raccordement :

$$\frac{I^{telecom}_{2012.3/2007}}{J^{telecom}_{2012.3/2010}} = \frac{0{,}833}{0{,}979} = 0{,}8509 \tag{3.58}$$

La valeur qu'aurait pris l'ancien indice au troisième trimestre de l'année 2013 s'il avait continué à être calculé est alors donnée par :

$$I'^{telecom}_{2013.3/2007} = J^{telecom}_{2013.3/2010} \times 0{,}8509 = 0{,}918 \times 0{,}8509 = 0{,}781 \tag{3.59}$$

Bien entendu, le résultat obtenu peut varier en fonction de la date k que l'on choisit pour procéder au raccordement. Il est important de souligner que l'opération de raccord d'indices ne conduit donc pas à un résultat certain au sens où (i) la composition des indices que l'on raccorde a en général évolué au cours du temps et (ii) d'un point de vue théorique, la propriété de circularité n'est pas vérifiée pour les indices synthétiques.

3.2 Indices chaînes

Ainsi que nous l'avons mentionné précédemment, les mutations économiques ont pour conséquence que des indices dont la base reste fixe sur longue période ne peuvent tenir compte de ces changements et ne sont en conséquence pas représentatifs de la réalité économique. Afin de pallier cette difficulté, il est possible de calculer des indices dont la base varie de date en date (ou de période en période). Cela consiste à généraliser le principe de raccord d'indices en définissant des **indices chaînes**.

Définition 3.6

L'**indice chaîne** $C^g_{t/0}$ calculé pour une grandeur g à la date t par rapport à la date de référence 0 s'écrit :

$$C^g_{t/0} = I^g_{t/t-1} \times C^g_{t-1/0} \tag{3.60}$$

soit encore :

$$C^g_{t/0} = I^g_{t/t-1} \times I^g_{t-1/t-2} \times \ldots \times I^g_{1/0} \tag{3.61}$$

Il est alors possible de définir un indice chaîne de Laspeyres (en remplaçant I^g par L^g dans l'équation (3.61)), un indice chaîne de Paasche (en remplaçant I^g par P^g dans

l'équation (3.61)) et un indice chaîne de Fisher (en remplaçant I^g par F^g dans l'équation (3.61)). Rappelons toutefois qu'il convient de prendre garde à l'interprétation au sens où un indice chaîne de Laspeyres ne constitue pas un indice de Laspeyres, le produit d'indices de Laspeyres ne donnant pas un indice de Laspeyres. Il en est de même pour les indices de Paasche et de Fisher.

4 Hétérogénéité et effet qualité

Dans les calculs d'indices que nous avons effectués jusqu'à présent, nous avons supposé implicitement que les biens ou produits considérés sont homogènes au sein d'une même classe. Ainsi, si l'on reprend l'exemple des fraises (▸ tableau 3.1), nous avons considéré que la classe des fraises était homogène. Or, en pratique, tel n'est évidemment pas le cas, cette classe est en effet hétérogène puisqu'il existe plus de 600 variétés de fraises (« Gariguette », « Charlotte », « Mara des bois », etc.). Ces variétés correspondent à différents niveaux de qualité, certaines variétés de fraises bénéficiant d'un signe d'identification de la qualité et de l'origine, comme le « label rouge », l'« Indication Géographique Protégée », etc. La prise en compte de l'hétérogénéité, et donc de la qualité, des produits n'est pas neutre quant au calcul et à l'interprétation des indices.

Afin d'illustrer ceci, considérons l'exemple de la production de vin en France. Répartissons cette classe en deux catégories i : les vins de consommation courante (dits « vins de France ») et les vins de qualité supérieure, cette dernière catégorie comprenant notamment les vins d'appellation d'origine protégée (AOP). Le tableau 3.4 reporte pour les années 2005 et 2013 et pour chacune des deux catégories considérées : le prix de vente d'un litre de vin (en euros par litre), la quantité produite (en millions de litres), ainsi que la valeur de la production définie comme le produit entre les prix et les quantités.

▼ **Tableau 3.4 Prix de vente (en euros par litre) et quantités produites (en millions de litres) de vin**

Variété i	2005			2013				
	p_0^i	q_0^i	V_0^i	p_t^i	q_t^i	V_t^i	$p_0^i q_t^i$	$p_t^i q_0^i$
Vin de consommation courante	2	575	1 150	2,5	415	1 037,5	830	1 437,5
Vin de qualité supérieure	3,5	210	735	3,8	348	1 322,4	1 218	798
Somme		785	1 885		763	2 359,9	2 048	2 235,5
Prix moyen	2,40			3,09				

On peut calculer les indices suivants pour le vin (g désignant le vin) :

- L'indice élémentaire du prix moyen du vin :

$$I^g_{t/0}(\bar{p}) = \frac{\bar{p}_t}{\bar{p}_0} = \frac{3{,}09}{2{,}40} = 1{,}288 \tag{3.62}$$

avec $\bar{p}_0 = \dfrac{\Sigma_i p_0^i q_0^i}{\Sigma_i q_0^i}$ et $\bar{p}_t = \dfrac{\Sigma_i p_t^i q_t^i}{\Sigma_i q_t^i}$.

- L'indice élémentaire de volume du vin :

$$I^g_{t/0}(q) = \frac{q_t}{q_0} = \frac{763}{785} = 0{,}972 \quad (3.63)$$

avec $q_0 = \Sigma_i q^i_0$ et $q_t = \Sigma_i q^i_t$.

- L'indice de valeur du vin :

$$I^V_{t/0} = \frac{V_t}{V_0} = \frac{2359{,}9}{1885} = 1{,}252 \quad (3.64)$$

avec $V_0 = \Sigma_i p^i_0 q^i_0$ et $V_t = \Sigma_i p^i_t q^i_t$.

Le calcul de ces indices nous montre que la valeur de la production de vin s'est accrue de 25,2 %, se décomposant en une diminution de la production de 2,8 % et une hausse du prix moyen de 28,8 %. Les calculs des indices de prix et de volume nécessitant toutefois d'agréger les prix et quantités de deux types de vin, il convient de calculer des indices synthétiques. Le tableau 3.5 reporte les valeurs des indices de Laspeyres et de Paasche calculées à partir des données figurant dans le tableau 3.4. On constate ainsi que la hausse du prix moyen du vin entre les deux années considérées se situe entre 15,2 % et 18,6 % selon l'indice retenu.

▼ **Tableau 3.5 Indices de Laspeyres et de Paasche**

	Laspeyres	Paasche
Prix ($\bar{p}$)	1,186	1,152
Quantité (q)	1,086	1,056

Si l'on recalcule la valeur de la production sur la base des indices de prix de Laspeyres et de Paasche, on obtient une hausse de la valeur entre 12 % (selon l'indice de Paasche) et 15,3 % (selon l'indice de Laspeyres), ce qui est bien différent de l'augmentation de 25,2 % précédemment mise en évidence. Comment expliquer une telle différence ? Cela provient d'une modification de la structure des ventes qui a évolué en faveur des vins de qualité supérieure. On peut en effet constater au regard du tableau 3.4 que la part dans la production totale du vin de qualité supérieure s'est accrue entre 2005 et 2013, passant de 26,8 % à 45,6 %. Cet accroissement de la qualité peut se quantifier à l'aide des indices de qualité (ou indices de structure).

Nous savons que :

$$I^V_{t/0} = P^g_{t/0}(q) \times L^g_{t/0}(\bar{p}) = P^g_{t/0}(\bar{p}) \times L^g_{t/0}(q) \quad (3.65)$$

et que :

$$I^V_{t/0} = \frac{V_t}{V_0} = \frac{\Sigma_i p^i_t q^i_t}{\Sigma_i p^i_0 q^i_0} = \frac{\Sigma_i q^i_t}{\Sigma_i q^i_0} \times \frac{\dfrac{\Sigma_i p^i_t q^i_t}{\Sigma_i q^i_t}}{\dfrac{\Sigma_i p^i_0 q^i_0}{\Sigma_i q^i_0}} = I^g_{t/0}(q) \times I^g_{t/0}(\bar{p}) \quad (3.66)$$

En égalisant les deux équations précédentes, il vient :

$$I^V_{t/0} = P^g_{t/0}(q) \times L^g_{t/0}(\bar{p}) = P^g_{t/0}(\bar{p}) \times L^g_{t/0}(q) = I^g_{t/0}(q) \times I^g_{t/0}(\bar{p}) \quad (3.67)$$

On en déduit donc les deux relations suivantes :

$$S_t = \frac{I^g_{t/0}(\bar{p})}{L^g_{t/0}(\bar{p})} = \frac{P^g_{t/0}(q)}{I^g_{t/0}(q)} \tag{3.68}$$

et :

$$S'_t = \frac{I^g_{t/0}(\bar{p})}{P^g_{t/0}(\bar{p})} = \frac{L^g_{t/0}(q)}{I^g_{t/0}(q)} \tag{3.69}$$

En détaillant les différentes formules, on peut réécrire les indices S_t et S'_t comme suit :

$$S_t = \frac{\dfrac{\Sigma_i p^i_t q^i_t}{\Sigma_i q^i_t}}{\dfrac{\Sigma_i p^i_t q^i_0}{\Sigma_i q^i_0}} \tag{3.70}$$

$$S'_t = \frac{\dfrac{\Sigma_i p^i_0 q^i_t}{\Sigma_i q^i_t}}{\dfrac{\Sigma_i p^i_0 q^i_0}{\Sigma_i q^i_0}} \tag{3.71}$$

On constate ainsi aisément que les indices S_t et S'_t décrivent les modifications apparues dans la composition de la grandeur (ou de la classe) étudiée entre la date de base et la date courante ; la différence entre les deux indices étant que le système des prix considéré est celui de la date courante pour S_t et de la date de base pour S'_t. Les indices S_t et S'_t sont appelés **indices de qualité** (ou **indices de structure**) et sont donc égaux à :

$$S_t = \frac{1{,}288}{1{,}186} = \frac{1{,}056}{0{,}972} = 1{,}086 \tag{3.72}$$

$$S'_t = \frac{1{,}288}{1{,}152} = \frac{1{,}086}{0{,}972} = 1{,}118 \tag{3.73}$$

En combinant les équations (3.67), (3.70) et (3.71), on obtient les égalités suivantes :

$$I^V_{t/0} = L^g_{t/0}(\bar{p}) \times S_t \times I^g_{t/0}(q) = P^g_{t/0}(\bar{p}) \times S'_t \times I^g_{t/0}(q) \tag{3.74}$$

Soit, avec nos données :

$$I^V_{t/0} = 1{,}186 \times 1{,}086 \times 0{,}972 = 1{,}152 \times 1{,}118 \times 0{,}972 = 1{,}252 \tag{3.75}$$

Cette dernière égalité nous permet ainsi de déduire que l'augmentation de 25,2 % de la valeur de la production de vin se décompose en :

– une hausse du prix moyen comprise entre 15,2 % (selon l'indice de Paasche) et 18,6 % (selon l'indice de Laspeyres) ;

– un accroissement de la qualité du vin comprise entre 8,6 % et 11,8 %, due à l'effet de structure, c'est-à-dire à la modification de la structure des ventes en faveur des vins de plus grande qualité ;

– une baisse de la quantité vendue de 2,8 %.



2 questions à

Axelle **Chauvet-Peyrard**

Chef de division à l'INSEE et précédemment responsable de la méthodologie de l'indice des prix à la consommation au sein de la direction des statistiques démographiques et sociales de l'INSEE



Comment l'indice des prix à la consommation mis à disposition par l'INSEE est-il calculé en pratique ?

L'indice des prix à la consommation (IPC) est un indicateur synthétique dont l'objectif est d'estimer la pression inflationniste à travers la mesure de l'évolution des prix des biens et services sur tout le territoire français. Le recensement complet de tous les produits offerts aux consommateurs étant matériellement impossible, cette mesure s'effectue par le biais d'un échantillonnage. Un « panier-type » de biens et services est ainsi défini et révisé chaque année de manière à être représentatif de la consommation réelle des ménages, telle qu'observée sur le territoire dans un passé récent. Parallèlement, on effectue un échantillonnage des points de vente au sein desquels seront observés tous les mois les prix des produits retenus dans le panier-type. Ces observations de prix sont ensuite agrégées sous forme d'indice. La formule retenue pour l'IPC comme pour l'IPCH (indice des prix à la consommation « harmonisé » au niveau européen) est, conformément à la réglementation européenne, celle de Laspeyres (voir équation (3.22)).

Quels problèmes peut poser un tel indice agrégé ?

Comme tout produit d'une agrégation, l'IPC reflète des réalités et des situations très diverses. Étant un indicateur d'inflation, il couvre l'intégralité du champ de la consommation des ménages, y compris des biens durables tels que les ordinateurs ou des services plus ou moins fréquemment consommés. Il y a souvent une mécompréhension de l'IPC, le grand public étant davantage attentif à ce qu'on appelle parfois le « prix du caddie ». De plus, pour être une mesure efficace de la pression inflationniste, il est nécessaire que la mesure d'évolution des prix se fasse *à qualité constante*. De cette façon, si le prix étiqueté d'un produit reste stable mais que sa qualité s'améliore, l'indice élémentaire du produit diminuera de fait. C'est ce qui se passe par exemple sur les ordinateurs. Et c'est également ce qui donne un IPC relativement stable depuis des années autour de 1,5 à 2 % d'inflation annuelle, alors que les ménages « vivent » une hausse des prix étiquetés jugée bien supérieure.

L'intégralité de l'entretien est disponible sur www.dunod.com. ■

Les points **clés**

- Un indice élémentaire est un nombre sans dimension permettant de résumer l'évolution d'une grandeur économique simple entre deux dates ou deux espaces différents à une même date, ou de comparer l'évolution de deux grandeurs simples.
- Les indices élémentaires vérifient les propriétés de circularité et de réversibilité.
- Un indice synthétique permet de résumer en une seule valeur l'information contenue dans plusieurs indices élémentaires.
- Les principaux indices synthétiques sont les indices de prix et de quantités de Laspeyres et de Paasche. Ces indices ne vérifient pas les propriétés de circularité et de réversibilité.
- L'indice des prix de Laspeyres décrit l'évolution du prix d'un panier de biens dont la composition est restée fixe entre les deux dates considérées.
- L'indice des prix de Paasche décrit l'évolution du prix d'un panier de biens dont la composition évolue en même temps que les prix.
- Le calcul d'indices de qualité permet de décomposer l'évolution de la valeur d'une grandeur en une partie due à la variation des prix, une partie due à la variation des quantités et une partie provenant de la modification de la structure de la grandeur étudiée due à l'effet qualité.

ÉVALUATION

► Corrigés sur www.dunod.com

QCM

Pour chacune des questions suivantes, indiquer si les affirmations sont vraies ou fausses (il peut y avoir plusieurs réponses vraies pour chaque question).

1 Propriétés sur les indices

a. Un indice élémentaire vérifie toujours les propriétés de circularité et de réversibilité.

b. Un indice vérifiant la propriété de circularité est tel que : $I^g_{t/0} \times I^g_{t'/0} = I^g_{t/t'}$.

c. Contrairement à l'indice de Paasche, l'indice de Laspeyres est réversible et satisfaisait la propriété de circularité.

d. L'indice de Fisher est réversible et ne satisfait pas la propriété d'agrégation.

e. L'indice de valeur est égal au produit des indices de prix et de quantités de Fisher.

2 Indices synthétiques

a. L'indice des prix de Paasche est la moyenne arithmétique des indices élémentaires, pondérée par les coefficients de la date de référence.

b. L'indice des prix de Paasche est la moyenne arithmétique des indices élémentaires, pondérée par les coefficients de la date courante.

c. Le calcul des indices des prix de Laspeyres et de Paasche ne nécessite pas que les indices élémentaires sur lesquels ils sont fondés soient basés à la même date.

d. Dans le cas de l'indice des prix de Paasche, la structure du panier de biens évolue avec les prix.

e. Dans le cas de l'indice des prix de Laspeyres, la structure du panier de biens est fixe.

3 On considère l'évolution du prix d'un quotidien de la presse écrite. On donne les indices élémentaires des prix suivants : $I_{2013/2000} = 1{,}30$ et $I_{2005/2000} = 1{,}50$. Que peut-on dire de l'évolution du prix du journal entre 2005 et 2013 ?

a. Le prix a augmenté de 86,7 %.

b. Le prix a augmenté de 13,3 %.

c. Le prix a baissé de 13,3 %.

d. Le prix a augmenté de 20 %.

e. Le prix a baissé de 20 %.

4 On donne l'indice élémentaire des prix suivant : $I_{2013/2005} = 1{,}283$. Que peut-on en déduire en termes d'évolution des prix entre 2005 et 2013 ?

a. Le prix du produit considéré a augmenté de 128,3 % entre 2005 et 2013.

b. Le prix du produit considéré a augmenté de 12,83 % entre 2005 et 2013.

c. Le prix du produit considéré a augmenté de 28,3 % entre 2005 et 2013.

d. Le prix du produit considéré a baissé de 71,7 % entre 2005 et 2013.

e. Le prix du produit considéré a augmenté de 71,7 % entre 2005 et 2013.

5 Le rapport suivant $\dfrac{\Sigma_i p^i_t q^i_t}{\Sigma_i p^i_t q^i_t \times \dfrac{p^i_0}{p^i_t}}$ correspond à la définition de :

a. L'indice des prix de Fisher.

b. L'indice des prix de Laspeyres.

c. L'indice des prix de Paasche.

d. L'indice des quantités de Laspeyres.

e. L'indice des quantités de Paasche.

Exercice

6 Indices élémentaires, synthétiques et effet qualité

Considérons un ménage consommant trois variétés de pommes : Golden, Pink Lady et Royal Gala. Le tableau 3.6 fournit, pour chacune de ces variétés, le prix

d'un kilogramme (en euros) ainsi que la quantité annuelle consommée (en kilogrammes).

▼ Tableau 3.6 Consommation de pommes d'un ménage

Variété i	Prix en 2010 p_0^i	Quantité en 2010 q_0^i	Prix en 2013 p_t^i	Quantité en 2013 q_t^i
Golden	2,80	10	3,10	15
Pink Lady	3,90	5	4,40	15
Royal Gala	2,10	15	2,20	10

1. Calculer le prix moyen du kilogramme de pommes ainsi que l'indice élémentaire du prix moyen.
2. Calculer l'indice élémentaire des quantités.
3. Calculer les indices de prix de Laspeyres et de Paasche de deux façons différentes.
4. Calculer les indices des quantités de Laspeyres et de Paasche.
5. Calculer les indices de prix et quantités de Fisher. La relation usuelle entre les indices synthétiques est-elle vérifiée ?
6. Peut-on conclure à une modification de la structure de consommation du ménage ? Ce résultat était-il attendu ?

Sujets d'examen

7 Université Paris Ouest, extrait

Un institut national de la statistique d'un pays X souhaite déterminer l'indice des prix de quatre groupes de biens A, B, C et D. Les enquêteurs fournissent les résultats donnés dans le tableau 3.7.

▼ Tableau 3.7 Prix et quantités des biens *A*, *B*, *C* et *D*

Bien h	Prix unitaire en 2000 p_0^h	Prix unitaire en 2014 p_t^h	Quantités consommées en 2000 q_0^h
A	169	610	210
B	81	265	220
C	1023	2470	30
D	32	64	470

L'institut a par ailleurs calculé que la consommation entre 2000 et 2014 a diminué de 23,8 % pour le groupe de biens A, augmenté de 4,54 % pour le groupe de biens B, diminué de 40 % pour le groupe de biens C et stagné pour le groupe de biens D.

1. Calculer les quantités consommées pour chacun des groupes de biens en 2014 (notées q_t^h).
2. Calculer les indices élémentaires des prix pour chaque groupe de biens, base 1 en 2000, notés $I_{t/0}^h$. Commenter.
3. Calculer l'indice des prix de Laspeyres (noté $L_{t/0}^h(p)$), l'indice des prix de Paasche (noté $P_{t/0}^h(p)$) et l'indice des prix de Fisher (noté $F_{t/0}^h(p)$). Tous ces indices sont calculés base 1 en 2000. Comparer les trois valeurs et commenter les résultats obtenus.

8 Université Paris Ouest, extrait

Une petite entreprise artisanale distribue trois types de produits ($h = a, b, c$). Le tableau 3.8 donne les prix (p_0^h) et quantités (q_0^h) vendues en 2010, ainsi que les indices élémentaires des prix ($i_{t/0}^h(p)$) et des quantités ($i_{t/0}^h(q)$) en 2014 (base 100 en 2010).

▼ Tableau 3.8 Prix, quantités et indices élémentaires – Biens *a*,*b*,*c*

h	p_0^h	q_0^h	$i_{t/0}^h(p)$	$i_{t/0}^h(q)$
a	4	5	200	45
b	8	5	75	100
c	10	4	120	125

1. Caractériser l'évolution des prix des trois produits sur la période considérée.
2. Déterminer les prix et les quantités des trois types de produits vendus en 2014.
3. Comparer les structures des chiffres d'affaires en 2010 et en 2014 selon les trois types de produits.
4. Calculer le nombre total de produits vendus en 2010 et en 2014, que l'on notera respectivement q_0 et q_t. En déduire l'indice élémentaire de la quantité totale de produits vendus (base 1 en 2010), noté $I_{t/0}(q)$. Commenter.
5. Sachant que l'indice de valeur $I_{t/0}^V$ est égal à 1,080 et en utilisant la valeur de $I_{t/0}(q)$ déterminée ci-dessus, calculer l'indice élémentaire du prix moyen des produits considérés. On donne par ailleurs le rapport suivant : $\frac{L_{t/0}(q)}{I_{t/0}(q)} = 1,133$. Commenter et interpréter ces résultats.

POUR ALLER PLUS LOIN

Indice des prix à la consommation et effet qualité

Nous avons ici traité de l'effet qualité lié à des variétés de biens différentes au sein d'une même classe. Notons que l'effet qualité apparaît également lors des mises à jour du panier de biens servant au calcul d'indices synthétiques, tel l'indice des prix à la consommation de l'INSEE. Les produits considérés dans le calcul de l'indice font l'objet de mises à jour régulières et il est donc possible que certains biens obsolètes disparaissent et soient éventuellement remplacés par des biens de qualité supérieure. Afin d'éviter que l'évolution du prix provienne d'une différence qualitative entre le bien remplacé et le bien remplaçant, l'INSEE élimine l'effet qualité en procédant à ce que l'on appelle un « ajustement de qualité ».

L'étude réalisée par Guédès (2004) montre que l'impact de ces ajustements de qualité n'est pas négligeable dans le calcul de l'indice des prix à la consommation. Ainsi, sur un nombre total de 450 000 produits, 14 000 ont été remplacés et deux tiers de ceux-ci ont fait l'objet d'un ajustement de qualité. Sur l'année 2003, Guédès (2004) met en évidence que ces ajustements ont réduit l'évolution de l'indice de 0,3 point de pourcentage : en l'absence de ces ajustements, l'indice des prix aurait augmenté de 2,5 % au lieu de 2,2 % sur l'année considérée. À un niveau plus fin d'analyse, l'impact des ajustements de qualité est différent selon les secteurs : il est plus fort pour les secteurs dans lesquels les produits sont rapidement renouvelés, tel le secteur de l'habillement, et plus faible dans le cas des produits alimentaires. Faut-il toujours cependant corriger de l'effet qualité ? Contrairement à ce que laisse penser l'exemple de l'indice des prix à la consommation, la réponse à cette question est négative. En effet, un même bien peut avoir plusieurs prix différents. Tel est par exemple le cas d'un bien agricole qui peut être soumis à des quotas différents sur certains segments et non soumis à quota sur un autre segment. Dans ce cas, il n'y a pas lieu de procéder à un ajustement de qualité et le prix à considérer est le prix moyen. Pour plus de détails sur le traitement des effets qualité, le lecteur intéressé pourra se reporter aux documents figurant sur le site de l'INSEE, dont Guédès (2004) et Berthier (2005).

Chapitre 4

Il est bien connu que les ventes de climatiseurs sont systématiquement plus élevées en été : l'étude de leur évolution au cours d'une année montre ainsi une hausse des ventes à l'approche de l'été, suivie d'une baisse à l'issue du mois d'août.

Supposons à présent que l'on cherche à analyser l'évolution des ventes non plus au cours d'une seule année, mais sur une période de temps plus longue, composée de plusieurs années consécutives, par exemple de 1980 à 2014. Sur l'ensemble des 35 années, les ventes ont-elles tendance à stagner, augmenter ou diminuer ? Les hausses et les baisses sont-elles régulières, c'est-à-dire ont-elles tendance à se répéter d'année en année ? Existe-t-il des fluctuations exceptionnelles ?

LES GRANDS AUTEURS



Warren M. Persons (1878-1937)

Warren M. Persons est un économiste statisticien. Professeur au Colorado College et remarqué pour ses travaux sur les baromètres économiques, il dirige le Committee on Economic Research crée en 1917 par l'Université de Harvard et édite le premier numéro de la célèbre *Review of Economic Statistics* en 1919.

Également connu pour ses travaux sur les indices, il a très largement contribué à l'analyse des séries temporelles en proposant une méthode complète de décomposition d'une série en quatre éléments : une composante tendancielle de long terme (séculaire), un mouvement cyclique, un mouvement saisonnier infra-annuel et des variations résiduelles ou accidentelles. Cette classification, dont les prémisses remontaient au début du XX[e] siècle, constitue depuis le **schéma de décomposition** de référence d'une série temporelle. ■

Séries temporelles : une introduction

Plan

Pré-requis

- **Savoir calculer** les caractéristiques d'une distribution (▸ chapitre 1).
- **Savoir lire et interpréter** un graphique dans le plan.
- **Connaître** l'ajustement par la droite des moindres carrés (▸ chapitre 2).

Objectifs

- **Analyser, décrire et expliquer** l'évolution d'un phénomène au cours du temps.
- **Repérer et identifier** la tendance, les variations saisonnières et les variations accidentelles dans l'évolution d'une variable au cours du temps.
- **Prévoir** le phénomène étudié à partir de son évolution passée.
- **Lisser** une série afin de faire ressortir son évolution générale.

Supposons par ailleurs qu'une grande enseigne vendant des climatiseurs ait mis en oeuvre une politique promotionnelle en juillet 2013. L'accroissement des ventes observé en juillet 2013 résulte-t-il de cette politique promotionnelle ou de la chaleur caractérisant la période ? En d'autres termes, la hausse des ventes en juillet 2013 est-elle plus importante que l'augmentation observée en juillet les années précédentes en l'absence de politique promotionnelle ? Comment déterminer la partie de la hausse résultant de l'élévation des températures et celle due à la promotion ? Pour répondre à l'ensemble de ces questions, il convient d'étudier l'évolution de la série des ventes de climatiseurs au cours du temps, ce que l'on appelle une **série temporelle**. Pour cela, il faut isoler les éléments constitutifs de l'évolution globale de la série temporelle, c'est-à-dire ses différentes composantes : sa tendance de long terme, ses variations saisonnières, ses variations accidentelles. L'objet de ce chapitre est ainsi de présenter les différents outils permettant de réaliser une telle analyse.

1 Exemples introductifs, définitions et description des séries temporelles

1.1 Quelques exemples

Considérons les figures suivantes reportant l'évolution de l'indice des prix à la consommation en France (▸ figure 4.1), du nombre de passagers sur les vols internationaux (▸ Aéroports de Paris, figure 4.2) et de l'indice boursier CAC 40 (▸ figure 4.3). Ces données constituent ce que l'on appelle des **séries temporelles** (ou **séries chronologiques** ou **chroniques**) au sens où une observation (c'est-à-dire une valeur) est observée à différentes dates espacées de façon régulière. Ainsi, la figure 4.2 reporte le nombre de passagers (en ordonnée) observé chaque mois entre janvier 1990 et janvier 2014 (en abscisse). De façon générale, on retient la définition suivante.

Définition 4.1

Une série temporelle Y_t est une suite d'observations $Y_1, Y_2, \ldots, Y_T$ ordonnées dans le temps, où t désigne le temps ($t = 1, \ldots, T$) et T est le nombre d'observations.

Le temps peut être une date ou une période, ce qui nous conduit à distinguer deux grands types de séries temporelles : série en **niveau** (ou **stock**) et série de **flux**. Un stock correspond ainsi à la valeur d'une variable à une date donnée, un flux se rapporte à l'évolution (augmentation ou diminution) de la variable entre deux dates. Dans le cas le plus fréquent, la durée entre deux dates (ou périodes de temps) successives d'observation des données est constante et correspond à ce que l'on appelle la **fréquence**. La fréquence peut être annuelle (▸ figure 4.1), trimestrielle, mensuelle (▸ figure 4.2), hebdomadaire, quotidienne (▸ figure 4.3), intra-quotidienne, etc.

FOCUS

La correction des jours ouvrés

Supposons que l'on souhaite comparer deux valeurs mensuelles d'une série temporelle. Les mois pouvant avoir des durées différentes (nombre de jours, de week-end, de jours fériés, etc.), il convient de tenir compte de ces différences lorsque l'on raisonne en termes de flux. Dans ce cas, on détermine les **séries corrigées des jours ouvrés (CJO)**. Prenons un exemple simple afin d'illustrer le problème. Considérons les exportations totales de biens, en millions d'euros, de la zone euro à 12 pays. Les données sont les suivantes (source : INSEE) :

- En décembre 2006, les exportations en volume s'élèvent à 124 164,9 millions d'euros. Le nombre de jours ouvrés (c'est-à-dire travaillés) est égal à 20.
- En janvier 2007, les exportations en volume s'élèvent à 121 815,4 millions d'euros. Le nombre de jours ouvrés est égal à 22.

Si l'on calcule la variation relative (c'est-à-dire le taux de croissance) du volume des exportations entre janvier 2007 et décembre 2006 sans tenir compte de la différence de durée entre les deux mois, on obtient :

$$\frac{121\,815{,}4 - 124164{,}9}{121\,815{,}4} = -0{,}0189$$

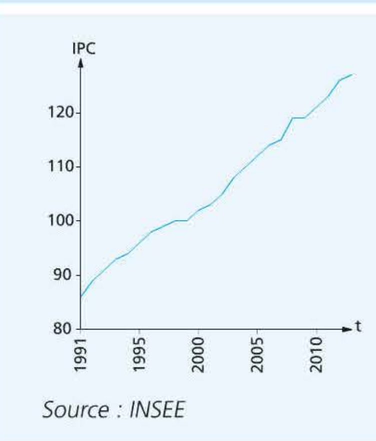
et l'on en déduit que les exportations ont légèrement diminué de 1,89 % d'un mois à l'autre. Un tel résultat est cependant faussé par le nombre différent de jours ouvrés entre les deux mois. Si l'on se fixe comme norme une durée de 21 jours ouvrés pour tous les mois de l'année, on peut calculer les exportations corrigées des jours ouvrés comme suit :

- En décembre 2006 :
$$\frac{124\,164{,}9}{20}\times 21 = 130\,373{,}145 \text{ millions d'euros.}$$
- En janvier 2007 :
$$\frac{121\,815{,}4}{22}\times 21 = 116\,278{,}336 \text{ millions d'euros.}$$

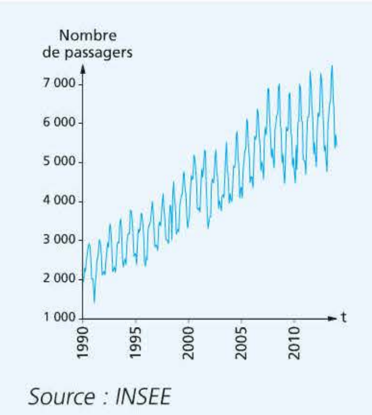
En calculant la variation relative entre janvier 2007 et décembre 2006, on constate que les exportations ont en fait diminué de façon plus importante, puisque la baisse s'élève à 10,81 %. De façon générale, on exprime la valeur CJO, Y_t^{CJO}, du flux Y_t au mois t comme suit :

$$Y_t^{CJO} = \frac{Y_t}{N_t^{JO}} \times D_{Ref} \qquad (4.1)$$

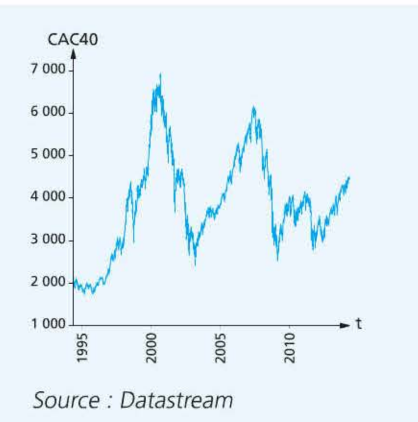
où N_t^{JO} désigne le nombre de jours ouvrés du mois t et D_{Ref} la durée moyenne de référence des mois à comparer.



▲ **Figure 4.1** Évolution de l'indice des prix à la consommation en France, données annuelles, 1990-2013 (base 100 en 1998)



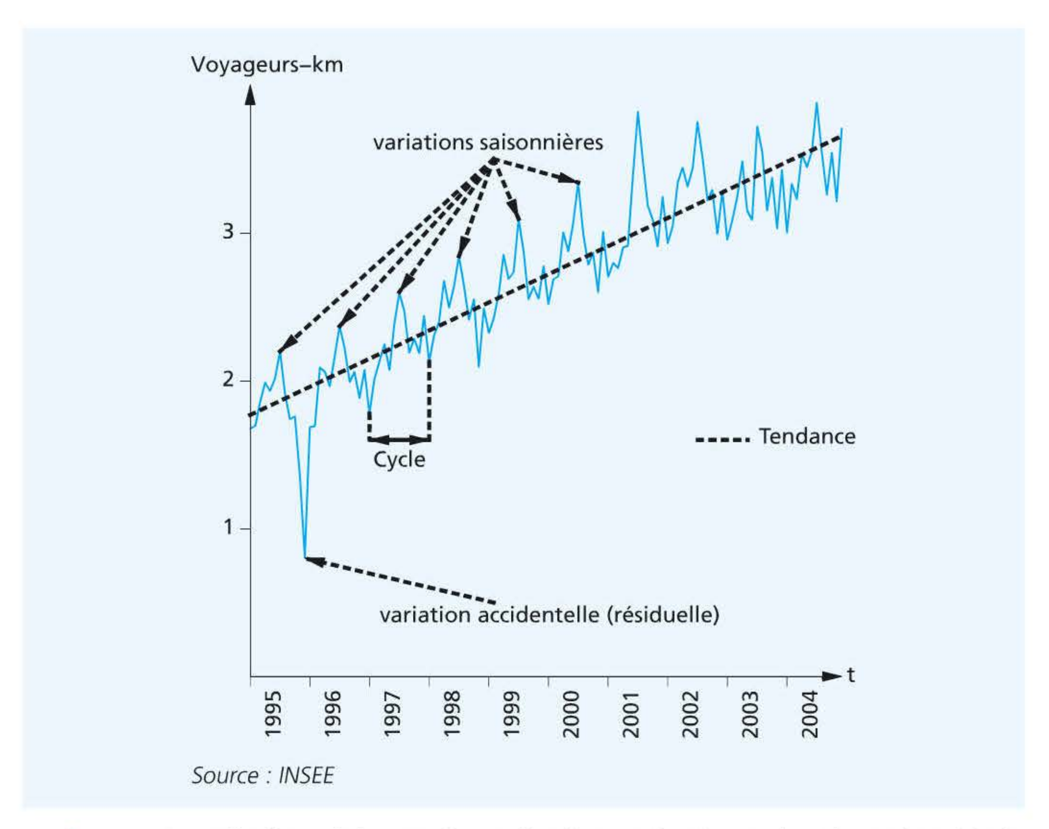
▲ **Figure 4.2** Évolution du nombre de passagers sur les vols internationaux (Aéroports de Paris), données mensuelles, janvier 1990-janvier 2014



▲ **Figure 4.3** Évolution de l'indice boursier CAC 40, données quotidiennes, 20/05/1994-20/05/2014

1.2 Description : les composantes d'une série temporelle

L'observation des figures 4.1 à 4.3 nous permet de faire ressortir un certain nombre de traits saillants. Au regard de la figure 4.1, il apparait que l'indice des prix à la consommation a tendance à augmenter au cours du temps. Il en est de même du trafic aérien (▶ figure 4.2), mais avec une particularité supplémentaire, à savoir la présence de fluctuations régulières, d'amplitude comparable, se répétant chaque année de façon périodique durant les mois de juillet et août. La figure 4.3 met quant à elle en évidence l'existence de fluctuations assez irrégulières, aléatoires de l'indice CAC 40. Les séries temporelles peuvent ainsi être décomposées en plusieurs éléments, appelés **composantes** (▶ figure 4.4) :



▲ **Figure 4.4** **Trafic ferroviaire sur les trains à grande vitesse (TGV) sur la période janvier 1995-décembre 2005, données mensuelles**

- La **tendance** (ou *trend* ou composante tendancielle ou composante séculaire), notée d_t, représente l'évolution à long terme de la série étudiée et reflète son comportement « normal », régulier. La figure 4.1 fait ainsi ressortir une tendance (linéaire) haussière de l'indice des prix à la consommation.
- Le **cycle** (ou composante cyclique) correspond à un mouvement décrivant des fluctuations autour de la tendance. En pratique, il est usuel en statistique de ne pas distinguer les composantes cyclique et tendancielle et de confondre l'évolution du

cycle avec celle de la tendance[1]. On considère donc dans la suite que la tendance regroupe à la fois la tendance et le cycle.

- La **composante saisonnière** (ou saisonnalité), notée s_t, traduit un phénomène se répétant à intervalles de temps réguliers (périodiques). Ce mouvement saisonnier est souvent considéré comme prévisible, tel est par exemple le cas des pics de trafic aérien observés tous les étés (▸ figure 4.2) correspondant aux départs en vacances. Les variations saisonnières se traduisent ainsi par des pics et des creux qui se répètent et reflètent les comportements (vacances, traditions, religions, etc.), les rythmes des saisons (tourisme, transport, consommation d'énergie, produits agricoles tels les fruits et légumes, etc.), ou encore d'autres facteurs économiques ou sociaux (soldes, etc.).
- La composante **résiduelle** (ou résidu ou aléa ou bruit), notée ϵ_t, correspond à des fluctuations irrégulières et aléatoires, comme cela est observé sur la figure 4.3. On intègre également dans cette composante les « phénomènes accidentels » comme des conditions météorologiques ou phénomènes climatiques exceptionnels (gel, sécheresse, inondations...), les grèves, les guerres, etc.

L'identification de ces différentes composantes constitue une étape cruciale afin de pouvoir décrire, expliquer et prévoir le phénomène étudié. À cette fin, on utilise des schémas de décomposition.

1.3 Décomposition d'une série temporelle

1.3.1 Les deux schémas classiques de décomposition

On distingue traditionnellement le schéma de décomposition additif et le schéma de décomposition multiplicatif :

- Selon le **schéma de décomposition additif**, la série Y_t s'écrit comme la somme des trois composantes, supposées indépendantes les unes des autres :

$$Y_t = d_t + s_t + \epsilon_t \qquad (4.2)$$

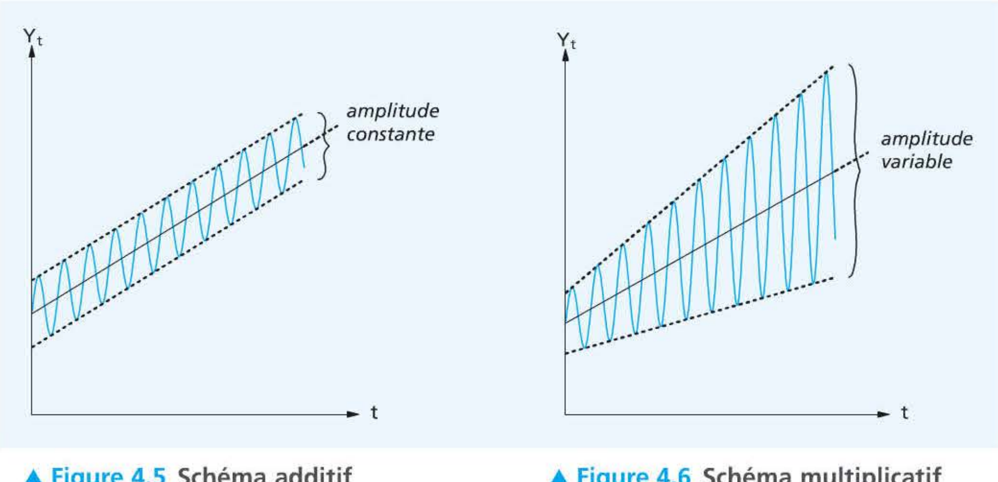
- Selon le **schéma de décomposition multiplicatif**, la série Y_t s'écrit comme suit[2] :

$$Y_t = d_t(1 + s_t)(1 + \epsilon_t) \qquad (4.3)$$

Dans le schéma de décomposition additif, en supposant que la composante résiduelle est faible, la différence entre la série et la tendance est sensiblement égale à la composante saisonnière. Tel n'est plus le cas pour le schéma multiplicatif puisque la différence entre la série et la tendance devient proportionnelle à la tendance. En d'autres termes, pour le schéma additif, l'amplitude de la composante saisonnière est constante au cours du temps (la saisonnalité est dite stable ou rigide), contrairement au cas du schéma multiplicatif dans lequel cette amplitude varie proportionnellement à la tendance au bruit près (la saisonnalité est dite évolutive). On peut en conséquence s'aider de graphiques pour effectuer le choix entre les deux types de schémas, ainsi que cela est représenté sur les figures 4.5 et 4.6.

1 L'une des raisons est liée au fait que les séries temporelles sont souvent trop courtes pour pouvoir procéder aisément à une telle décomposition.

2 Notons qu'il est possible de définir plusieurs types de schémas multiplicatifs – comme par exemple $Y_t = d_t(1 + s_t) + \epsilon_t$ – celui retenu ici (équation (4.3)) est le plus utilisé en économie.



▲ Figure 4.5 Schéma additif

▲ Figure 4.6 Schéma multiplicatif

Remarque : Supposons, à titre d'exemple, que l'on dispose de données trimestrielles. Une autre technique visant à choisir entre les deux schémas de décomposition consiste à calculer pour chaque année la moyenne et l'écart-type des 4 observations de l'année. On reporte alors graphiquement (ou dans un tableau, dit **tableau de Buys-Ballot**) les valeurs de la moyenne (en abscisse) en fonction des valeurs de l'écart-type (en ordonnée). S'il ressort graphiquement une absence de lien entre la moyenne et l'écart-type (droite parallèle à l'axe des abscisses), il convient de retenir un schéma additif. Si tel n'est pas le cas, on choisit un modèle multiplicatif.

FOCUS

La composante saisonnière

Dans le cas de la composante saisonnière, l'étendue des intervalles auxquels se répète le phénomène étudié est constante et appelée **période de la saisonnalité** (notée P).

À titre d'exemple, dans le cas d'une série mensuelle (respectivement trimestrielle) pour laquelle une saisonnalité est systématiquement observée le même mois (respectivement trimestre) chaque année, on a $P = 12$ (respectivement $P = 4$).

La constance généralement supposée de la composante saisonnière sur chaque période P, soit $s_t = s_{t+P} = s_{t+2P} = \ldots$, implique que l'effet de la composante saisonnière est en moyenne nul sur la période P : $\Sigma_{j=1}^{P} s_j = 0$.

Cette contrainte provient du **principe de conservation des aires** selon lequel l'influence des variations saisonnières est neutre sur la période P : les variations saisonnières se compensent sur la période P au sens où la somme des aires entre la série et la tendance situées au dessus de la tendance est égale à celle située en dessous de la tendance. Les valeurs s_j, $j = 1,\ldots,P$, prises par s_t sont appelées **coefficients saisonniers**.

Ainsi, si l'on considère le nombre d'entrées quotidiennes au cinéma les valeurs $s_1 = -0,5$ et $s_6 = 0,8$ signifient que la fréquentation des salles de cinéma diminue de 50 % le lundi et augmente de 80 % le samedi par rapport à l'ensemble de la semaine.

1.3.2 Synthèse : démarche générale pour l'analyse d'une série temporelle à partir de ses composantes

Afin de décrire l'évolution d'une série temporelle, il est nécessaire d'étudier ses différentes composantes. Pour cela, on procède en quatre étapes :

- **Étape 1** : on choisit le schéma – additif ou multiplicatif – de décomposition en s'aidant, comme nous l'avons vu précédemment, de graphiques. On étudie ensuite séparément la tendance (étape 2) et la composante saisonnière (étape 3).
- **Étape 2** : détermination de la tendance. Cette étape consiste à isoler la tendance afin d'étudier l'évolution de long terme de la série.
- **Étape 3** : correction des variations saisonnières et/ou des jours ouvrés. Cette étape consiste à corriger la tendance des variations saisonnières dans le cas d'une série en niveau, et des variations saisonnières et des jours ouvrés dans le cas d'une série en flux. Considérons par exemple les ventes mensuelles de jouets au sein d'un grand magasin et supposons que ce dernier a mis en place une politique promotionnelle au mois de décembre 2013. L'étude de la série corrigée des variations saisonnières (CVS) permettra ainsi de déterminer si le pic observé en décembre 2013 est plus important que les pics observés en décembre les autres années et résulte de l'effet de la politique promotionnelle mise en place. De façon plus générale, afin de pouvoir comparer les ventes d'un mois sur l'autre, il est nécessaire de tenir compte de l'effet de la saisonnalité en supprimant la composante saisonnière, tel est précisément l'objet de la correction des variations saisonnières.
- **Étape 4** : une fois l'estimation de la tendance et la détermination de la série CVS effectuées, il est possible en ôtant les valeurs de la tendance à celles de la série CVS de faire apparaître les seules variations de la série dues à la composante résiduelle.

2 Détermination et estimation de la tendance

Afin de simplifier la présentation, considérons le cas de séries temporelles pour lesquelles il est possible de faire abstraction de la composante saisonnière et composées, en conséquence, uniquement d'une tendance et d'une composante résiduelle. Notre objectif est ici d'identifier, et donc d'extraire, la tendance d'une telle série afin de procéder ensuite à son estimation. À cette fin, on cherche à lisser la série temporelle pour n'en conserver que sa tendance. Pour cela, on utilise des techniques de lissage comme les **moyennes mobiles** et les méthodes de **lissage exponentiel**.

2.1 Les moyennes mobiles

Afin de comprendre intuitivement la technique des moyennes mobiles, considérons le cas d'une série temporelle dont la tendance d_t est linéaire[3], soit $d_t = at + b$. On a donc

3 Nous ne considérons dans cet ouvrage que le cas de tendances linéaires. Il existe bien entendu d'autres types de tendances, plus complexes, comme les tendances polynomiales de degré supérieur à 1, exponentielles, logistiques, hyperboliques, etc.

$Y_t = at+b+\epsilon_t$. L'objectif est ici d'estimer la tendance, c'est-à-dire d'obtenir des valeurs $\hat{a}$ et $\hat{b}$ (dites valeurs estimées) pour les « vrais » coefficients a (pente) et b (ordonnée à l'origine). Ainsi que nous l'avons vu dans l'analyse de régression (► chapitre 2), l'application de la méthode des moindres carrés ordinaires (MCO) nous fournit de telles estimations et nous donne pour l'estimateur du coefficient de pente : $\hat{a} = \frac{Cov(t,Y_t)}{V(t)}$. En procédant de la sorte, on suppose que l'estimation de la tendance est la même sur l'ensemble de la période et l'on tient compte de l'ensemble des observations (T) antérieures et postérieures à la date courante t, en leur attribuant le même poids ($1/T$). Or, il peut être préférable d'attribuer un poids différent aux observations – en accordant, par exemple, un poids plus faible aux observations plus lointaines – ou de ne retenir qu'une partie des observations. La méthode des moyennes mobiles consiste ainsi à retenir en t un nombre fixé N d'observations les plus récentes et à négliger les observations les plus anciennes. À chaque date t, l'échantillon d'observations N se modifie donc, conduisant à une série d'estimations de la tendance.

Plus précisément, le principe des moyennes mobiles consiste à remplacer N observations consécutives par leur moyenne arithmétique, en « faisant glisser » ce calcul de date en date. N est un nombre entier, appelé **ordre** (ou **longueur**) de la moyenne mobile[4].

FOCUS

Choix de l'ordre d'une moyenne mobile

Le choix de l'ordre N de la moyenne mobile dépend de l'ampleur des fluctuations de la série Y_t. Si l'amplitude des fluctuations est restreinte, une faible valeur de N suffit à lisser la série. En revanche, si cette amplitude est importante, il convient de retenir une valeur élevée de N pour lisser la série. Il est ainsi possible de s'aider de graphiques pour effectuer ce choix. En pratique, on a coutume de retenir les valeurs suivantes :

- Pour des données annuelles : $N = 3$ ou $N = 5$.
- Pour des données trimestrielles : $N = 4$.
- Pour des données mensuelles : $N = 12$.

Rappelons que nous avons supposé dans cette section l'absence de composante saisonnière. Si la série étudiée comporte une composante saisonnière et que l'on souhaite distinguer la tendance de cette composante, il convient d'appliquer à cette série une moyenne mobile dont l'ordre est égal à la période P de la saisonnalité ou est un multiple de P (à titre d'exemple, pour des données mensuelles, $P = 12$). Dans ce cas, les variations saisonnières sont éliminées.

Cela résulte du principe de conservation des aires appliqué à la composante saisonnière : sur une période donnée, la partie positive et la partie négative de cette composante se compensent, conduisant à une valeur nulle à l'issue de l'application de la moyenne mobile.

4 On parle également parfois de taille de la fenêtre de la moyenne mobile.

Exemple

Prenons le cas d'une série Y_t, pour $t = 1,2,\ldots,7$, et déterminons sa moyenne mobile M_t d'ordre 3. Les résultats figurent dans le tableau 4.1. Notons que pour $t = 1$ et $t = 7$ la moyenne mobile d'ordre 3 ne peut être calculée et l'on perd ainsi une valeur à chaque extrémité de la série. La moyenne mobile M_t calculée dans la troisième colonne du tableau 4.1 est une moyenne mobile centrée : le nombre d'observations pris en compte avant la date à laquelle est calculée la moyenne mobile est égal au nombre d'observations pris en compte après cette même date. Il existe également des moyennes mobiles dites simples ou non centrées (notées MS_t) que l'on peut calculer comme indiqué dans la dernière colonne du tableau 4.1. La moyenne mobile simple d'ordre $N = 3$ ne peut pas être calculée pour $t = 1$ et $t = 2$ car $t < N$. En pratique, les moyennes mobiles simples ne sont adaptées qu'au cas de tendances constantes, soit $d_t = b$, et ne peuvent être utilisées en cas de tendances plus « complexes » (linéaires, quadratiques, exponentielles, ...).

▼ **Tableau 4.1** **Principe de calcul d'une moyenne mobile d'ordre 3**

t	Y_t	M_t	MS_t
1	Y_1		
2	Y_2	$M_2 = (Y_1 + Y_2 + Y_3)/3$	
3	Y_3	$M_3 = (Y_2 + Y_3 + Y_4)/3$	$MS_3 = (Y_1 + Y_2 + Y_3)/3$
4	Y_4	$M_4 = (Y_3 + Y_4 + Y_5)/3$	$MS_4 = (Y_2 + Y_3 + Y_4)/3$
5	Y_5	$M_5 = (Y_4 + Y_5 + Y_6)/3$	$MS_5 = (Y_3 + Y_4 + Y_5)/3$
6	Y_6	$M_6 = (Y_5 + Y_6 + Y_7)/3$	$MS_6 = (Y_4 + Y_5 + Y_6)/3$
7	Y_7		$MS_7 = (Y_5 + Y_6 + Y_7)/3$

On distingue les moyennes mobiles d'ordre impair des moyennes mobiles d'ordre pair :

- Cas où N est impair : $N = 2k + 1$ (avec $k = 1,2,\ldots$).

Définition 4.2

On appelle **opérateur de lissage par moyennes mobiles d'ordre impair** $N = 2k + 1$, l'opérateur M défini pour $t = k + 1,\ldots,T - k$ qui transforme la série Y_t en une série M_t telle que :

$$M_t = \frac{Y_{t-k} + Y_{t-k+1} + \ldots + Y_t + \ldots + Y_{t+k-1} + Y_{t+k}}{2k + 1} \tag{4.4}$$

- Cas où N est pair : $N = 2k$ (avec $k = 1,2,\ldots$).

Définition 4.3

On appelle **opérateur de lissage par moyennes mobiles d'ordre pair** $N = 2k$, l'opérateur M défini pour $t = k + 1,\ldots,T - k$ qui transforme la série Y_t en une série M_t telle que :

$$M_t = \frac{\frac{1}{2}Y_{t-k} + Y_{t-k+1} + \ldots + Y_t + \ldots + Y_{t+k-1} + \frac{1}{2}Y_{t+k}}{2k} \tag{4.5}$$

Notons que dans les deux cas, lorsque l'on calcule une moyenne mobile, on perd k valeurs à chaque extrémité de la série Y_t : la série M_t comprend donc moins d'observations que la série Y_t.

L'estimation de la tendance d_t par moyennes mobiles est alors immédiate puisque cela consiste à choisir M_t comme estimateur, soit $\hat{d}_t = M_t$. On constate ainsi que l'estimateur de la tendance n'est plus constant (comme dans le cas des MCO) mais varie avec t. La différence $u_t = d_t - \hat{d}_t$ désigne l'**erreur de prévision** de la tendance commise en t. L'amplitude de l'erreur diminue avec la valeur de N, c'est pour cela que l'on parle de lissage de la série Y_t.

On déduit des développements précédents que la **prévision** $\hat{Y}_{T+h}$ faite en T de la série Y pour la date $T+h$ où h désigne l'horizon de prévision ($h = 1,2,\ldots$) est donnée par :

$$\hat{Y}_{T+h} = \hat{d}_{T+h} = \hat{a}(T+h) + \hat{b} = \hat{a}T + \hat{b} + \hat{a}h = \hat{d}_T + \hat{a}h \quad (4.6)$$

Naturellement, si la tendance est constante $d_t = b$, on a plus simplement : $\hat{Y}_{T+h} = \hat{d}_T = M_T$.

2.2 Le lissage exponentiel simple

Le **lissage exponentiel simple (LES)** est une technique s'appliquant au cas de séries pour lesquelles la tendance est constante au cours du temps, soit : $d_t = b$ et donc $Y_t = b + \epsilon_t$[5]. Le principe du LES consiste à estimer en t la tendance d'une série en considérant l'ensemble des observations antérieures à t[6], mais en accordant un poids de plus en plus faible aux observations de plus en plus lointaines et, en conséquence, en donnant plus d'importance aux observations récentes.

Définition 4.4

On appelle opérateur de lissage exponentiel simple (LES) de paramètre α la fonction L qui transforme la série temporelle Y_t en une série L_t telle que :

$$L_t = \alpha Y_t + (1-\alpha)L_{t-1} \quad (4.7)$$

où $0 < \alpha < 1$ est appelé **paramètre (ou constante) de lissage** et $t = 1,\ldots,T$.

En procédant par récurrence, on peut réécrire l'équation (4.7) comme suit :

$$L_t = \alpha Y_t + (1-\alpha)(\alpha Y_{t-1} + (1-\alpha)L_{t-2}) = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + (1-\alpha)^2 L_{t-2} \quad (4.8)$$

En poursuivant ainsi de suite, on obtient la formule développée du LES exprimant la série lissée comme une combinaison linéaire de l'ensemble des valeurs de la série Y_t :

$$L_t = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \ldots + \alpha(1-\alpha)^{t-1}Y_1 + (1-\alpha)^t L_0 \quad (4.9)$$

où L_0 désigne la condition initiale du LES. Cette dernière expression illustre bien le fait que le poids des observations passées diminue au fur et à mesure que l'on s'éloigne dans le temps. En particulier, le poids $(1-\alpha)^t$ associé à L_0 diminue rapidement vers 0

5 Dans les méthodes de lissage exponentiel, on suppose que ϵ_t est d'espérance nulle, de variance constante et non autocorrélé. On dit alors que ϵ_t est un **bruit blanc**.

6 Les techniques de lissage exponentiel (simple et double) se distinguent ainsi des méthodes de moyennes mobiles par le fait qu'elles tiennent compte de toutes les observations antérieures, et pas seulement des N dernières observations.

lorsque t augmente, traduisant le fait que la condition initiale joue très rapidement un rôle négligeable dans le calcul de L_t. Notons qu'en pratique, on retient fréquemment comme valeur initiale soit la moyenne de la série Y_t, soit la première observation Y_1.

On en déduit immédiatement l'estimation de la tendance, qui s'exprime comme une moyenne arithmétique pondérée de l'observation de la série en t (Y_t) et de la tendance estimée en $t-1$ ($\hat{d}_{t-1}$) :

$$\hat{d}_t = L_t = \alpha Y_t + (1-\alpha)\hat{d}_{t-1} \tag{4.10}$$

avec $\hat{d}_0 = L_0$.

On peut encore écrire cette équation comme suit :

$$\hat{d}_t = \hat{d}_{t-1} + \alpha(Y_t - \hat{d}_{t-1}) = \hat{d}_{t-1} + \alpha u_t \tag{4.11}$$

où u_t désigne l'erreur de prévision réalisée en t :

$$u_t = Y_t - \hat{d}_{t-1} = Y_t - L_{t-1} \tag{4.12}$$

L'estimation de la tendance en t apparaît donc comme la correction de l'estimation réalisée en $t-1$ d'une fraction α de l'erreur de prévision en t.

La prévision de la série réalisée à la date T pour la date $T+h$, où h désigne l'horizon de prévision, est naturellement donnée par la dernière estimation de la tendance[7] :

$$\hat{Y}_{T+h} = \hat{d}_T = L_T \tag{4.13}$$

FOCUS

Choix de la valeur du paramètre de lissage α

Les pondérations associées aux valeurs de Y_t diminuent de façon exponentielle au cours du temps : plus le paramètre de lissage α est élevé, plus la décroissance est rapide. Ainsi, plus α est proche de 0 (respectivement de 1), plus on tient compte des observations lointaines (respectivement récentes). En conséquence, une valeur élevée pour α permet une adaptation plus rapide à un changement de niveau de la série.

En pratique, il est possible de s'aider de graphiques : si la série semble peu « heurtée », une faible valeur de α (inférieure ou égale à 0,3) peut être suffisante pour lisser la série.

Outre cette méthode graphique, on peut aussi recourir à des critères statistiques : le paramètre α du LES est ainsi souvent choisit de sorte à minimiser la somme des carrés des erreurs de prévision $\Sigma_t u_t^2$.

2.3 Le lissage exponentiel double

Dans le cas où la tendance n'est plus constante au cours du temps, il convient de ne plus utiliser le LES[8] et d'avoir recours au **lissage exponentiel double (LED)**. Considérons ainsi une tendance linéaire $d_t = at + b$. Le LED consiste à appliquer le LES à la série lissée L_t, où $L_t = \alpha Y_t + (1-\alpha)L_{t-1}$.

7 Notons que dans le cas du LES où la tendance est constante au cours du temps, la prévision effectuée à la date T est une valeur constante, indépendante de l'horizon h, on a donc : $\hat{Y}_{T+h} = \hat{Y}_T$.

8 Ainsi que nous le démontrons dans l'encadré « Pour aller plus loin », il existe en effet un biais systématique entre la valeur observée et la valeur lissée lorsque l'on utilise le LES pour estimer la tendance dans le cas où cette dernière n'est pas constante au cours du temps.

POUR ALLER PLUS LOIN
► Voir p. 105

Définition 4.5

On appelle opérateur de lissage exponentiel double (LED) de paramètre α la fonction LL qui transforme la série temporelle L_t en une série LL_t telle que :

$$LL_t = \alpha L_t + (1 - \alpha)LL_{t-1} \quad (4.14)$$

où $0 < \alpha < 1$ et $t = 1,\dots,T$.

POUR ALLER PLUS LOIN
► Voir p. 105

En procédant par récurrence, on montre que l'estimation de la tendance par le biais du LED est donnée par les équations suivantes :

$$\hat{d}_t = 2L_t - LL_t \quad (4.15)$$

l'estimation $\hat{a}_t$ de la pente à la date t s'écrivant :

$$\hat{a}_t = \frac{\alpha}{1-\alpha}(L_t - LL_t) \quad (4.16)$$

La prévision de la série réalisée à la date T pour la date $T + h$, où h désigne l'horizon de prévision, est alors donnée par :

$$\hat{Y}_{T+h} = \hat{a}_T(T + h) + \hat{b} = \hat{d}_T + \hat{a}_T h = 2L_T - LL_T + h\frac{\alpha}{1-\alpha}(L_T - LL_T) \quad (4.17)$$

3 Désaisonnalisation : la correction des variations saisonnières

Ainsi que nous l'avons précédemment mentionné, l'estimation de la tendance d'une série ne doit pas être perturbée par la présence de variations saisonnières, ces dernières masquant en partie la tendance générale de la série. Si la série étudiée présente une composante saisonnière, il convient donc de la purger de ses variations saisonnières en calculant une nouvelle série appelée **série corrigée des variations saisonnières (CVS)** ou **série désaisonnalisée**.

3.1 Principe général

Les variations saisonnières constituent des écarts à la tendance dans le cas d'un schéma de décomposition additif et des rapports à la tendance pour un schéma multiplicatif.

- Cas du schéma de décomposition additif : $Y_t = d_t + s_t + \epsilon_t$. Supposons que l'on dispose d'une estimation $\hat{s}_t$ de la composante saisonnière s_t. La série corrigée des variations saisonnières, notée Y_t^{CVS}, est donnée par :

$$Y_t^{CVS} = Y_t - \hat{s}_t \quad (4.18)$$

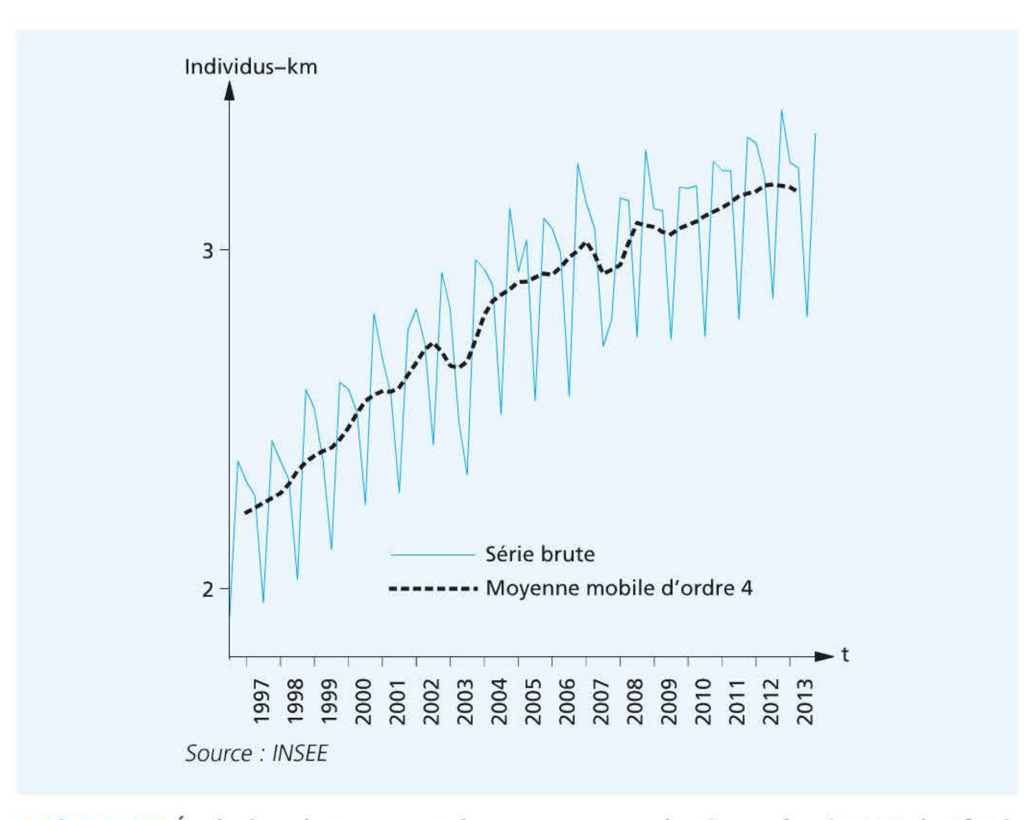
- Cas du schéma de décomposition multiplicatif : $Y_t = d_t(1 + s_t)(1 + \epsilon_t) = d_t S_t(1 + \epsilon_t)$, avec $S_t = 1 + s_t$. Supposons que l'on dispose d'une estimation $\hat{S}_t$ de la composante

saisonnière S_t. La série corrigée des variations saisonnières, notée Y_t^{CVS}, est donnée par :

$$Y_t^{CVS} = \frac{Y_t}{\hat{S}_t} \tag{4.19}$$

3.2 Calcul pas à pas d'une série CVS

Afin d'illustrer le calcul d'une série CVS, considérons la série de transport des voyageurs sur le réseau ferré de la RATP (source : INSEE). La série, dont les premières et dernières valeurs sont reportées dans le tableau 4.2, est à fréquence trimestrielle et couvre la période allant du troisième trimestre de l'année 1996 au dernier trimestre 2013. Comme on peut le constater sur la figure 4.7, cette série présente une saisonnalité puisque l'on observe de façon systématique une baisse au troisième trimestre de chaque année, pouvant s'expliquer par le fait que le réseau RATP est moins emprunté durant la période des vacances d'été. S'agissant du choix du schéma de décomposition (additif ou multiplicatif), l'observation de l'amplitude de la composante saisonnière nous conduit plutôt à opter pour un schéma additif. À des fins pédagogiques, nous considérons toutefois les deux schémas par la suite.



▲ **Figure 4.7** **Évolution du transport de voyageurs sur le réseau ferré RATP du 3e trimestre 1996 au 4e trimestre 2013 (en individus-kilomètres)**

3.2.1 Étape 1 : estimation de la tendance

Commençons par estimer la tendance d_t de la série de trafic RATP Y_t. La série étant trimestrielle, nous considérons une moyenne mobile d'ordre 4. Le tableau 4.2 reporte les valeurs de la série ainsi lissée, $\hat{d}_t$. La figure 4.7 montre que la série exhibe une tendance à la hausse sur l'ensemble de la période et que les variations saisonnières ont bien été éliminées par l'application d'une moyenne mobile dont l'ordre est égal à la période de la saisonnalité ($N = P = 4$).

▼ **Tableau 4.2** **Transport de voyageurs sur le réseau ferré RATP (en individus-kilomètres)**

Date	Y_t	$\hat{d}_t$	z_{ij} Additif	z_{ij} Multiplicatif
1996/3	1,912			
1996/4	2,377			
1997/1	2,316	2,226	0,090	1,041
1997/2	2,275	2,239	0,036	1,016
1997/3	1,958	2,254	−0,296	0,869
1997/4	2,437	2,268	0,169	1,075
1998/1	2,378	2,283	0,095	1,042
1998/2	2,323	2,310	0,013	1,006
...	...	...	...	...
2012/1	3,316	3,173	0,143	1,045
2012/2	3,218	3,191	0,027	1,009
2012/3	2,856	3,194	−0,338	0,894
2012/4	3,414	3,190	0,224	1,070
2013/1	3,259	3,186	0,073	1,023
2013/2	3,243	3,171	0,072	1,023
2013/3	2,801			
2013/4	3,345			

Source : INSEE.

3.2.2 Étape 2 : estimation et interprétation des coefficients saisonniers

Notons Y_{ij} l'observation du j^e trimestre de l'année i, avec $i = 1,\ldots,I$ et $j = 1,2,3,4$ où I désigne le nombre total d'années considérées dans l'échantillon. On peut donc écrire les schémas de décomposition comme suit[9] :

- Schéma de décomposition additif : $Y_{ij} = d_{ij} + s_j + \epsilon_{ij}$

9 Le mouvement saisonnier se répétant à l'identique d'année en année, il ne dépend pas de i, expliquant le fait que s soit uniquement indicé par j.

- Schéma de décomposition multiplicatif : $Y_{ij} = d_{ij}(1 + s_j)(1 + \epsilon_{ij}) = d_{ij}S_j(1 + \epsilon_{ij})$, avec $S_j = 1 + s_j$

On calcule les écarts (schéma additif) et les rapports (schéma multiplicatif) à la tendance estimée comme suit :

- Pour le schéma additif : $z_{ij} = Y_{ij} - \hat{d}_{ij}$
- Pour le schéma multiplicatif : $z_{ij} = \dfrac{Y_{ij}}{\hat{d}_{ij}}$

Les coefficients ainsi calculés sont reportés dans le tableau 4.2 pour les deux schémas. On estime ensuite les coefficients saisonniers $\hat{s}_j$ en effectuant la moyenne des valeurs z_{ij} pour chacun des trimestres sur l'ensemble des années considérées, soit[10] :

$$\hat{s}_j = \frac{1}{I}\Sigma_{i=1}^{I} z_{ij} \tag{4.20}$$

Les résultats obtenus, pour les deux schémas de décomposition sont reportés dans le tableau 4.3. Rappelons que l'on doit avoir $\Sigma_{j=1}^{4}\hat{s}_j = 0$, soit $\bar{s} = 0$ pour le schéma additif et $\Sigma_{j=1}^{4}\hat{S}_j = 4$, soit $\bar{S} = 1$ pour le schéma multiplicatif, $\bar{s}$ et $\bar{S}$ représentant la moyenne sur les quatre trimestres de $\hat{s}_j$ et $\hat{S}_j$ respectivement. Si ces contraintes ne sont pas vérifiées, il convient de corriger les coefficients saisonniers en calculant les coefficients normalisés $\hat{s}'_j = \hat{s}_j - \bar{s}$ et $\hat{S}'_j = \hat{S}_j - \bar{S}$. Nous avons ici :

$$\bar{s} = \frac{0{,}118 + 0{,}036 - 0{,}327 + 0{,}175}{4} = 0{,}0005 \tag{4.21}$$

et

$$\bar{S} = \frac{1{,}043 + 1{,}012 + 0{,}883 + 1{,}063}{4} = 1{,}0001 \tag{4.22}$$

On constate que ces valeurs sont très proches des valeurs attendues, bien que différant très légèrement de 0 et 1. La dernière ligne du tableau donne en conséquence les valeurs normalisées des coefficients saisonniers.

Si l'on considère le schéma additif, un coefficient négatif témoigne d'une valeur de la série inférieure à la tendance, alors qu'un coefficient positif illustre une valeur de la série supérieure à la tendance. Ainsi, la valeur estimée du coefficient saisonnier au troisième trimestre est égale à $\hat{s}_3 = -0{,}328$, ce qui signifie qu'au troisième trimestre le trafic sur le réseau est systématiquement inférieur à la tendance (de 0,33 individus-kilomètres environ). Au contraire, au dernier trimestre, $\hat{s}_4 = 0{,}174$, illustrant le fait que le trafic est supérieur à la tendance. On peut donc s'attendre chaque année à une hausse du nombre de voyageurs sur le réseau RATP entre le troisième et le dernier trimestre.

De façon similaire, les calculs pour le schéma multiplicatif nous montrent que le coefficient saisonnier $\hat{S}_3$ est inférieur à 1 ($\hat{S}_3 = 0{,}882$) : tous les ans, au troisième trimestre le trafic RATP est inférieur à la tendance de 11,8 % environ. En revanche, au dernier trimestre, le nombre de voyageurs transportés par la RATP est supérieur à la tendance, d'environ 6,3 %. Là encore, la RATP peut donc prévoir, tous les ans, une augmentation du nombre de voyageurs entre le troisième et le dernier trimestre.

10 Notons qu'il est également possible de remplacer la moyenne par la médiane des valeurs z_{ij} afin d'éviter l'effet des valeurs extrêmes.

▼ Tableau 4.3 Calcul des coefficients saisonniers

	Schéma additif				Schéma multiplicatif			
Date	Trim. 1	Trim. 2	Trim. 3	Trim. 4	Trim. 1	Trim. 2	Trim. 3	Trim. 4
1997	0,090	0,036	−0,296	0,169	1,041	1,016	0,869	1,075
1998	0,095	0,013	−0,322	0,213	1,042	1,006	0,863	1,090
1999	0,140	−0,026	−0,303	0,169	1,059	0,989	0,875	1,069
2000	0,114	0,004	−0,306	0,242	1,046	1,002	0,880	1,094
2001	0,095	−0,006	−0,311	0,135	1,037	0,998	0,880	1,051
2002	0,158	0,019	−0,304	0,236	1,059	1,007	0,889	1,088
2003	0,167	−0,158	−0,336	0,236	1,063	0,940	0,874	1,086
2004	0,134	0,046	−0,354	0,240	1,048	1,016	0,876	1,083
2005	0,029	0,123	−0,365	0,163	1,010	1,042	0,875	1,056
2006	0,136	0,042	−0,412	0,258	1,046	1,014	0,862	1,086
2007	0,116	0,077	−0,214	−0,148	1,038	1,026	0,927	0,950
2008	0,198	0,124	−0,339	0,223	1,067	1,041	0,890	1,073
2009	0,054	0,063	−0,313	0,122	1,017	1,021	0,897	1,040
2010	0,108	0,105	−0,357	0,149	1,035	1,034	0,885	1,048
2011	0,110	0,094	−0,364	0,166	1,035	1,030	0,885	1,052
2012	0,143	0,027	−0,338	0,224	1,045	1,009	0,894	1,070
Moyenne	0,118	0,036	−0,327	0,175	1,043	1,012	0,883	1,063
Coef. normalisés	0,117	0,036	−0,328	0,174	1,043	1,012	0,882	1,063

Note : « Moyenne » est la moyenne des coefficients $\hat{s}_j$ pour le schéma additif et des coefficients $\hat{S}_j$ pour le schéma multiplicatif.

3.2.3 Étape 3 : calcul de la série CVS

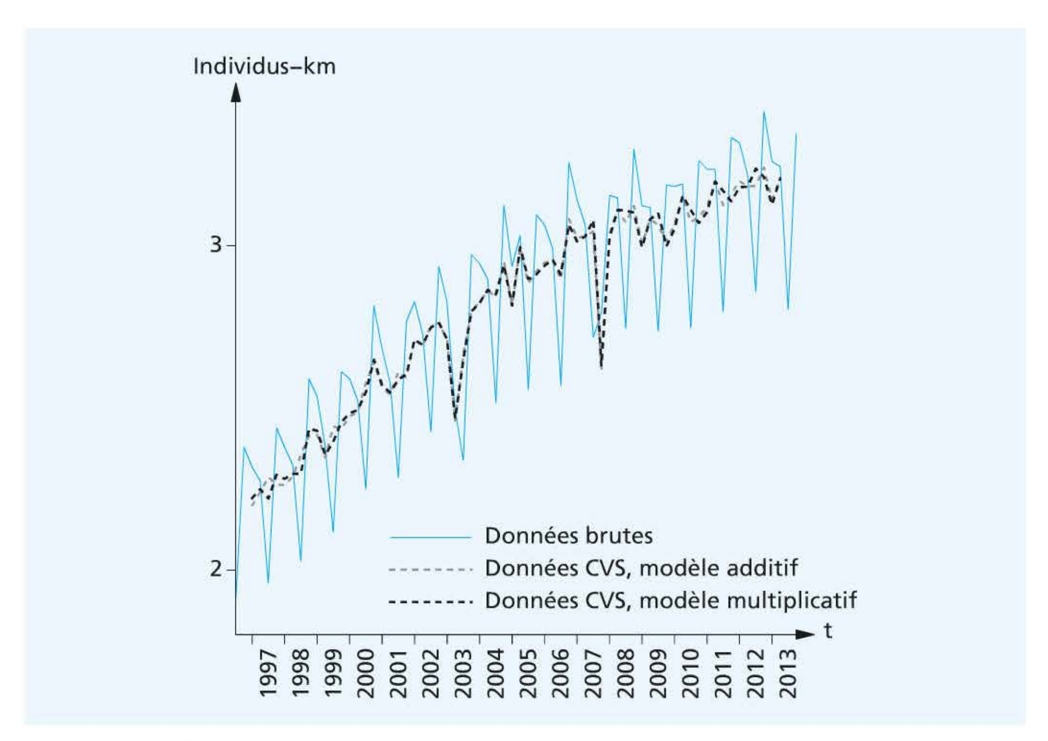
Disposant à présent d'une estimation de la tendance (étape 1) et de la saisonnalité (étape 2), on en déduit directement la série CVS :

- Pour le schéma additif : $Y_{ij}^{CVS} = Y_{ij} - \hat{s}_j$ soit $Y_t^{CVS} = Y_t - \hat{s}_t$.
- Pour le schéma multiplicatif : $Y_{ij}^{CVS} = \dfrac{Y_{ij}}{\hat{S}_j}$ soit $Y_t^{CVS} = \dfrac{Y_t}{\hat{S}_t}$.

Les séries CVS ainsi obtenues selon les deux schémas sont reportées sur la figure 4.8. On constate que la série a été lissée de sa composante saisonnière : Y_t^{CVS} représente la façon dont aurait évolué Y_t en l'absence de phénomène saisonnier. On relève en particulier l'existence de deux évolutions particulièrement marquées au deuxième trimestre 2003 et au dernier trimestre 2007 correspondant à d'importants épisodes de grèves (variations accidentelles). En pratique, on poursuit la procédure en procédant à un nouveau lissage sur la série CVS afin d'estimer sa tendance (étape 4). Cette dernière servira alors de base à la prévision des valeurs désaisonnalisées de la série. Doté de cette nouvelle estimation de la tendance, on peut effectuer une nouvelle estimation de la saisonnalité (étape 5) dans le but de chercher à améliorer l'estimation de la

composante saisonnière. Ces deux étapes 4 et 5 peuvent être réitérées. Une fois la tendance et la composante saisonnière estimées, il est possible d'effectuer des prévisions de la série. Ainsi, en notant h l'horizon de prévision ($h \geq 1$) et en considérant à titre d'exemple le schéma additif, on a :

$$\hat{Y}_{T+h} = \hat{d}_{T+h} + \hat{s}_j \tag{4.23}$$



▲ **Figure 4.8** **Évolution du transport de voyageurs sur le réseau ferré RATP du 3e trimestre 1996 au 4e trimestre 2013 (en individus-kilomètres) : série brute et séries CVS**

Les points **clés**

→ Une série temporelle peut être décomposée en trois éléments : une tendance, une composante saisonnière et une composante résiduelle.

→ La tendance représente le comportement de long terme de la série étudiée.

→ La méthode des moyennes mobiles permet de lisser une série temporelle afin d'identifier sa tendance.

→ La désaisonnalisation consiste à corriger une série des variations saisonnières.

→ Les méthodes de lissage exponentiel permettent de prévoir une série à court terme sur la base de ses seules observations passées.



1 question à

Laurent **Ferrara**

Chef du service de Macroéconomie Internationale à la Banque de France et Professeur associé à l'Université Paris Ouest





Dans le cadre de vos études et recherches à la Banque de France, est-il important d'identifier la tendance d'une série macroéconomique ou financière ?

Le service de macroéconomie internationale de la Banque de France s'occupe du suivi des pays industrialisés n'appartenant pas à la zone euro et de sujets plus transversaux comme les taux de change, les matières premières ou les déséquilibres mondiaux. Lorsque nous faisons le suivi conjoncturel de l'économie des pays ou des marchés financiers pour le gouverneur de la Banque de France, il est important de dégager des messages clairs et de ne pas focaliser uniquement sur les derniers chiffres qui peuvent refléter des événements exceptionnels, voire inexpliqués. Dans un langage statistique et économétrique, cela signifie que nous cherchons à identifier le signal (c'est-à-dire la tendance et/ou le cycle) par rapport au bruit (ou composante résiduelle).

L'extraction du signal est essentielle pour l'analyse économique car les relations macroéconomiques théoriques entre les variables sur lesquelles nous nous appuyons concernent en général le moyen terme, voire le long terme. Ainsi, filtrer le bruit de très court terme et récupérer les tendances de moyen et long termes constituent une grande partie du travail de conjoncturiste. Dans cette optique, les méthodes statistiques et économétriques sont d'une grande utilité, notamment les méthodes de lissage de type moyenne mobile ou les techniques de filtrage qui permettent de décomposer les variables macroéconomiques entre une tendance de long terme et une composante cyclique.

L'intégralité de l'entretien est disponible sur www.dunod.com. ■

ÉVALUATION

► Corrigés sur www.dunod.com

QCM

Pour chacune des questions suivantes, indiquer si les affirmations sont vraies ou fausses (il peut y avoir plusieurs réponses vraies pour chaque question).

1 Dans un schéma de décomposition additif, lorsque l'on ôte à la série brute Y_t les coefficients saisonniers, on obtient :

a. La série estimée $\hat{Y}_t$.

b. Une estimation de la tendance.

c. La composante résiduelle.

d. La série corrigée des variations saisonnières.

e. La série corrigée des jours ouvrés.

2 Dans le cas d'un schéma de décomposition additif :

a. La moyenne des coefficients saisonniers est égale à l'unité.

b. La moyenne des coefficients saisonniers est égale à la période P de la saisonnalité.

c. La somme des coefficients saisonniers est nulle.

d. La somme des coefficients saisonniers est égale à l'unité.

e. La moyenne des coefficients saisonniers est nulle.

3 Principe de conservation des aires et moyennes mobiles

a. En accord avec le principe de conservation des aires, l'influence des variations saisonnières est neutre sur l'ensemble de la période étudiée.

b. En accord avec le principe de conservation des aires, l'influence des variations saisonnières est neutre sur la période de la saisonnalité.

c. Une moyenne mobile d'ordre P permet d'éliminer la saisonnalité d'ordre P d'une série temporelle.

d. Une moyenne mobile consiste à lisser une série en tenant compte systématiquement de l'ensemble de ses observations.

e. Dans une moyenne mobile d'ordre N, N observations consécutives sont remplacées par leur moyenne arithmétique.

4 Lissage exponentiel

a. Le LES est une technique pouvant s'appliquer aux séries caractérisées par tout type de tendance.

b. Le LED s'applique au cas de séries présentant une tendance linéaire ainsi qu'une composante saisonnière.

c. Plus le paramètre de lissage du LES est proche de 0, plus le poids des observations récentes est important.

d. Lorsque la série comporte une composante saisonnière, on peut utiliser indifféremment le LES ou le LED.

e. Dans le cas du LES, la valeur prévue de la série ne dépend pas de l'horizon de prévision.

5 Les valeurs de l'indice Euro Stoxx sont égales à 3093,124 en janvier 2014 et 3085,865 en février de la même année. Les mois de janvier et de février comprennent respectivement 23 et 20 jours ouvrés. Si l'on retient 21 comme nombre moyen de jours ouvrés pour l'ensemble des mois de l'année, entre janvier et février 2004 l'indice Euro Stoxx a :

a. diminué de 0,235 %.

b. diminué de 23,5 %.

c. diminué de 14,73 %.

d. augmenté de 14,73 %.

e. n'a pas évolué.

Exercices

6 Étude de l'évolution de l'indice Euro Stoxx 50

On s'intéresse à l'évolution du cours de l'indice Euro Stoxx 50 lors du mois d'avril 2014. On dispose à cette fin des données quotidiennes du 1/04/2014 au 30/04/2014 (tableau 4.4).

1. Appliquer la méthode du LES à cette série pour les trois valeurs suivantes du paramètre de lissage α : 0,2 ; 0,5 et 0,74.
2. Déterminer la valeur prévue en t, $\hat{Y}_t$, de la série pour chacune des valeurs de α. En déduire les valeurs de l'erreur de prévision u_t et calculer les sommes des carrés des erreurs de prévision relatives à chaque valeur de α.
3. Selon les résultats obtenus à la question précédente, quelle est parmi les trois valeurs de α proposées celle qui vous semble la plus appropriée ? Ce résultat était-il attendu ?
4. Quelle est la valeur prévue de la série Y_t pour le 1/05/2014 ? Même question pour le 2/05/2014.

▼ Tableau 4.4 Indice Euro Stoxx 50 (Y_t), avril 2014

Date	Y_t	Date	Y_t
01/04/2014	3186,336	16/04/2014	3139,264
02/04/2014	3187,45	17/04/2014	3155,806
03/04/2014	3206,759	18/04/2014	3155,806
04/04/2014	3230,332	21/04/2014	3155,806
07/04/2014	3185,967	22/04/2014	3199,686
08/04/2014	3177,658	23/04/2014	3175,973
09/04/2014	3182,793	24/04/2014	3189,809
10/04/2014	3152,864	25/04/2014	3147,397
11/04/2014	3116,54	28/04/2014	3165,837
14/04/2014	3131,566	29/04/2014	3208,685
15/04/2014	3091,524	30/04/2014	3198,387

Source : Datastream.

7 Étude des ventes de voitures neuves en France

On considère la série trimestrielle des ventes de voitures neuves en France entre le premier trimestre de l'année 2006 et le quatrième trimestre de l'année 2013 (tableau 4.5).

1. Réorganiser les données figurant dans le tableau 4.5 sous la forme d'un tableau à double entrée présentant en ligne les années et en colonne les trimestres.
2. À partir du tableau précédemment construit, peut-on mettre en évidence l'existence d'une saisonnalité dans la série ? Commenter.
3. Calculer la moyenne et l'écart-type, année par année.
4. On s'intéresse à l'existence possible d'un lien entre les valeurs annuelles de la moyenne et celles de l'écart-type. Déterminer le coefficient de la pente de la droite de régression de l'écart-type sur la moyenne et calculer le coefficient de corrélation entre les deux séries. Que peut-on en déduire quant à la nature du schéma de décomposition à adopter ?

▼ Tableau 4.5 Ventes de voitures neuves (Y_t), France

Date	Y_t	Date	Y_t
2006/1	526502	2010/1	590869
2006/2	582727	2010/2	597383
2006/3	409138	2010/3	432029
2006/4	482182	2010/4	589905
2007/1	519191	2011/1	642627
2007/2	561417	2011/2	557981
2007/3	442753	2011/3	422728
2007/4	541182	2011/4	537592
2008/1	526121	2012/1	501879
2008/2	602775	2012/2	520836
2008/3	446869	2012/3	370006
2008/4	474518	2012/4	464292
2009/1	505456	2013/1	431388
2009/2	625859	2013/2	480730
2009/3	482186	2013/3	367112
2009/4	655170	2013/4	477721

Source : ministère de l'Écologie, du Développement durable et de l'Énergie.

POUR ALLER PLUS LOIN

Biais du LES et équations du LED

Commençons par montrer que l'application du LES à une série comportant une tendance non constante de la forme $d_t = at + b$ conduit à un biais systématique. Partons de l'équation du LES $L_t = \alpha Y_t + (1-\alpha)L_{t-1}$ et raisonnons par récurrence en se donnant une condition initiale $L_0 = a + b$ (avec $L_0 = Y_1$) :

$$L_1 = \alpha Y_1 + (1-\alpha)L_0 = \alpha(a+b) + (1-\alpha)(a+b) = a+b \quad (4.24)$$

$$L_2 = \alpha Y_2 + (1-\alpha)L_1 = \alpha(2a+b) + (1-\alpha)(a+b) = 2a + b - a(1-\alpha) \quad (4.25)$$

$$L_3 = \alpha Y_3 + (1-\alpha)L_2 = \alpha(3a+b) + (1-\alpha)(2a+b) - a(1-\alpha)^2 = 3a + b - a[(1-\alpha) + (1-\alpha)^2] \quad (4.26)$$

On en déduit :

$$L_t = (at+b) - a[(1-\alpha) + (1-\alpha)^2 + \ldots + (1-\alpha)^{t-1}] \quad (4.27)$$

soit encore :

$$L_t = d_t - a(1-\alpha)[1 + (1-\alpha) + (1-\alpha)^2 + \ldots + (1-\alpha)^{t-2}] \quad (4.28)$$

D'où :

$$L_t = d_t - a\frac{1-\alpha}{\alpha}[1 - (1-\alpha)^{t-1}] \quad (4.29)$$

On constate qu'il existe un **biais**, c'est-à-dire un écart entre la valeur estimée $L_t = \hat{d}_t$ de la tendance et sa valeur observée d_t, systématique égal à $-a\frac{1-\alpha}{\alpha}[1-(1-\alpha)^{t-1}]$. Appliquer le LES à une série comportant une tendance non constante au cours du temps conduit donc à une estimation biaisée de la tendance.

Pour des valeurs élevées de t, $(1-\alpha)^{t-1}$ tend vers zéro et l'on peut écrire :

$$L_t = \hat{d}_t = d_t - a\frac{1-\alpha}{\alpha} = (at+b) - a\frac{1-\alpha}{\alpha} \quad (4.30)$$

et le biais asymptotique (c'est-à-dire pour des valeurs élevées de t) est ainsi donné par $-a\frac{1-\alpha}{\alpha}$.

Afin de démontrer les équations (4.15) et (4.16) du LED, il suffit de remplacer $d_t = at + b$ par $d_t = at + b - a\frac{1-\alpha}{\alpha}$ dans la relation (4.30) :

$$LL_t = L_t - a\frac{1-\alpha}{\alpha} = d_t - 2a\frac{1-\alpha}{\alpha} \quad (4.31)$$

D'où, en effectuant la différence entre les équations (4.30) et (4.31) :

$$L_t - LL_t = d_t - a\frac{1-\alpha}{\alpha} - L_t + a\frac{1-\alpha}{\alpha} \quad (4.32)$$

Soit finalement :

$$2L_t - LL_t = d_t \quad (4.33)$$

On peut en outre réécrire l'équation (4.31) comme suit :

$$L_t - LL_t = a\frac{1-\alpha}{\alpha} \quad (4.34)$$

D'où l'on tire :

$$a = \frac{\alpha}{1-\alpha}(L_t - LL_t) \quad (4.35)$$

On en déduit alors immédiatement les valeurs estimées reportées dans les équations (4.15) et (4.16) du LED.

Partie 2

Probabilités et variable aléatoire

Les concepts de probabilité et de variable aléatoire constituent les notions fondamentales de l'analyse statistique. La théorie moderne des probabilités repose sur la notion d'espace probabilisable et définit la probabilité comme une mesure appliquée sur une tribu d'événements de cet espace.

La notion de variable aléatoire se comprend alors comme une application mesurable, c'est-à-dire une sorte de fonction, définie d'un univers probabilisé vers un univers des réalisations probabilisables.

Ainsi, toute la théorie moderne des statistiques et ses applications dans le domaine de l'entreprise et de la vie courante reposent sur ces deux notions.

Chapitre 5

Les notions de risque et de probabilité sont omniprésentes dans le monde économique. Un exemple parmi tant d'autres est celui de la détection de la fraude fiscale. Une des fraudes les plus coûteuses pour les finances publiques est celle dite du « carrousel » qui est un système sophistiqué permettant à des individus de récupérer les trop-perçus de TVA *via* la création puis la suppression rapide de sociétés fictives dans plusieurs pays de l'Union européenne. On estime qu'en France cette fraude coûterait à l'État près de 13 milliards d'euros par an. Le problème c'est qu'il est humainement impossible de contrôler les millions de déclarations de récupération de TVA.

Une solution consiste à modéliser statistiquement ces transactions et à orienter les contrôles vers les transactions présentant la plus grande probabilité d'être frauduleuses. Chaque transaction est alors associée à une probabilité d'être frauduleuse dépendant des caractéristiques de la société, de la nature de la transaction, etc. Une solution de ce type a aidé l'administration fiscale belge à récupérer plus d'un milliard d'euros en quelques années.

LES GRANDS AUTEURS



Andreï Kolmogorov (1903-1987)

Andreï Kolmogorov est un mathématicien russe considéré comme l'un des pères fondateurs de la théorie des probabilités.
Après des études à l'Université de Moscou, Kolmogorov publie ses premiers travaux concernant la théorie des ensembles et l'analyse de Fourier dans les années 1920.
Mais c'est dans son manuel, *Fondements de la théorie des probabilités*, publié en allemand en 1933, qu'il formalise la notion de σ-algèbre et pose les bases de l'axiomatisation du calcul des probabilités que nous verrons dans ce chapitre. ■

Probabilités

Plan

Pré-requis

→ **Connaître** les principales notions d'analyse combinatoire.

Objectifs

→ **Connaître** les notions d'expérience aléatoire et d'événement.
→ **Comprendre** la notion d'univers probabilisable.
→ **Comprendre** la notion de probabilité.
→ **Comprendre** la notion de probabilité conditionnelle.
→ **Connaître** la notion d'indépendance.
→ **Savoir appliquer** le théorème de Bayes et la formule des probabilités totales.

De façon très globale, la statistique peut être définie comme la branche des mathématiques consacrée à la modélisation du **risque**. Or, la notion de risque est associée à celle de **probabilité**. Mais qu'est-ce qu'une probabilité ? Dans le langage courant, une probabilité est souvent comprise comme une sorte de mesure du caractère probable d'un événement. En fait, on confond souvent les notions de probabilité et de fréquence. Ainsi, lorsque l'on prononce la phrase « un accident de la route sur deux est dû à l'alcool ou à la vitesse excessive », fait-on référence ou non à une probabilité ?

C'est pourquoi des mathématiciens, comme Andreï Kolmogorov (▸ encadré les grands auteurs), se sont efforcés de définir précisément la notion de probabilité dans le cadre de ce que l'on appelle aujourd'hui la **théorie des probabilités**. Cette théorie permet de définir la probabilité comme une **mesure** appliquée à une **tribu d'événements**. Cette formalisation est essentielle car elle fonde la notion de variable aléatoire (▸ chapitre 6) utilisée dans de très nombreux domaines d'application tels que le marketing quantitatif, les mathématiques financières, le traitement d'image, etc.

1 Définitions

1.1 Expérience aléatoire et univers des possibles

Définition 5.1

Une **expérience aléatoire** est une expérience renouvelable, en théorie ou en pratique, et qui, renouvelée dans des conditions identiques ne donne pas forcément le même résultat à chaque renouvellement.

L'exemple typique d'une expérience aléatoire renouvelable en pratique est celle du lancer de pièce. Il est possible de répéter plusieurs fois un lancer de pièce dans les mêmes conditions : à chaque lancer, on n'obtiendra pas nécessairement le même résultat, *i.e.* « pile » ou « face ». Mais d'autres expériences aléatoires ne peuvent pas être renouvelées en pratique dans les mêmes conditions. Si l'on adopte une vision purement aléatoire du monde, on peut par exemple considérer que la réussite d'un étudiant au baccalauréat est une expérience aléatoire qui peut aboutir à l'un des deux résultats « admis » ou « non admis ». Bien évidemment, il s'agit dans ce cas d'une représentation théorique de la réussite à l'examen. En effet, dans la pratique, cette expérience aléatoire ne peut pas être reproduite plusieurs fois dans les mêmes conditions, c'est-à-dire la même année, avec le même sujet et le même niveau de préparation et de maturité de l'étudiant, etc.

Définition 5.2

L'**univers des possibles** (ou univers), noté Ω (prononcer *grand omega*), est défini par l'ensemble de tous les résultats possibles qui peuvent être obtenus au cours d'une expérience aléatoire.

On distingue les univers comprenant un nombre *fini* de résultats de ceux comprenant un nombre *infini* de résultats. Parmi les univers infinis, on distingue les univers *infinis non dénombrables* des univers *infinis dénombrables*. Par exemple, l'univers $\Omega = \{\omega_1, \ldots, \omega_n, ..\} = \{\omega_i, i \in \mathbb{N}\}$ est un univers infini dénombrable puisque l'on peut identifier chacun des éléments de Ω, même s'il en existe une infinité. En revanche, $\Omega = \mathbb{R}$ ou $\Omega =]-\infty, a]$ sont des exemples d'univers infinis non dénombrables. Dans le cas d'un univers fini ou infini dénombrable, la taille de l'univers est appelée **cardinalité** et est représentée par l'opérateur $card\,(\Omega)$.

Exemple

On considère une expérience aléatoire correspondant au lancer d'un dé à 6 faces. L'univers (fini) des possibles est alors défini par :

$$\Omega = \{1,2,3,4,5,6\} \tag{5.1}$$

La cardinalité de cet univers est égale à 6, *i.e.* $card\,(\Omega) = 6$.

Exemple

On admet que le nombre de gouttes d'eau qui tombent pendant une durée d'une heure sur une surface donnée est le résultat d'une expérience aléatoire théorique. L'univers des possibles est alors défini par l'ensemble des valeurs entières (car on ne peut pas compter 1/2 goutte), positives ou nulles (car on ne peut pas avoir un nombre de gouttes négatif) :

$$\Omega = \{0,1,2,3,\ldots,n\} \tag{5.2}$$

La cardinalité de cet univers est égale à n, *i.e.* $card\,(\Omega) = n$. Si $n \in \mathbb{N}$, cet univers est fini. Si, au contraire $n = \infty$ et $\Omega = \mathbb{N}$, cet univers est infini, mais dénombrable.

Remarque : On oppose la théorie des **probabilités discrètes**, fondée sur un univers fini ou infini dénombrable, et la théorie des **probabilités continues**, fondée sur un univers infini non dénombrable.

Dans la suite de ce chapitre, nous nous focaliserons essentiellement sur le cas d'un univers des possibles fini ou infini dénombrable (théorie des probabilités discrètes). Nous étendrons les résultats obtenus au cas infini non dénombrable (théorie des probabilités continues).

1.2 Événements

De façon générale, à partir d'un ensemble, il est toujours possible de définir des sous-ensembles. Il en va de même pour un univers des possibles. Un sous-ensemble de l'univers des possibles est appelé une **partie** ou un **événement**.

Définition 5.3

Un **événement** (ou une partie) A est un sous-ensemble de l'univers des possibles Ω, vérifiant[1] $A \subset \Omega$. Un événement constitué d'un seul élément, *i.e.* pour lequel $card\,(A) = 1$, est un **événement élémentaire** (ou singleton).

1 Le symbole $\subset$ signifie « est inclus dans ».

Exemple

On considère une expérience aléatoire correspondant au lancer d'un dé à 6 faces, telle que $\Omega = \{1,2,3,4,5,6\}$. L'événement « nombre pair », noté A, correspond au sous-ensemble de l'univers Ω défini par $A = \{2,4,6\}$. L'événement « nombre impair », noté B, correspond au sous-ensemble $B = \{1,3,5\}$. On peut en outre définir un autre événement C, sans lui attribuer un nom particulier, tel que $C = \{1,5\}$ par exemple. L'événement $D = \{1\}$ est un événement élémentaire ou singleton.

À partir de la définition d'un événement, nous pouvons à présent introduire les notions d'événement certain et d'événement impossible.

Définition 5.4

Un **événement certain** correspond à l'univers des possibles Ω.

Pour bien comprendre le concept d'événement certain, considérons une expérience aléatoire particulière où l'univers des possibles se ramène à un seul événement élémentaire. Par exemple, si l'on reprend notre exemple de la réussite au baccalauréat, supposons que $\Omega = \{$« admis »$\}$. Dans ce cas, il n'y a qu'un seul résultat possible : notre étudiant est donc sûr de réussir son examen. Il s'agit d'un événement certain. De façon générale, l'« événement » Ω, quelle que soit sa cardinalité, est un événement certain. À l'inverse, on peut définir un événement impossible.

Définition 5.5

Un **événement impossible** est un événement qui ne se réalise jamais. Il correspond à l'ensemble vide, noté $\varnothing$.

Par exemple, l'événement « avoir 30/20 de moyenne au baccalauréat » est un événement impossible. On le représente donc par $\varnothing$.

Remarque : Il est important de noter que l'événement impossible est un ensemble vide, mais qu'un ensemble vide reste un ensemble. Il s'agit d'une sorte de boîte vide, mais d'une boîte. Par convention, cet ensemble (ou sous-ensemble) fait toujours partie de l'univers des possibles, *i.e.* $\varnothing \subset \Omega$. Par exemple, les deux notations $\Omega = \{\omega_1,\ldots,\omega_n\}$ et $\Omega = \{\omega_1,\ldots,\omega_n,\varnothing\}$ sont équivalentes.

Nous pouvons à présent combiner des événements à l'aide d'*opérations assemblistes* (pour utiliser le vocabulaire de la théorie des ensembles).

Définition 5.6

Soient deux événements A et B. La réalisation de l'événement C, défini par $C = A \cup B$ (lire A **union** B), implique la réalisation de l'événement A ou de l'événement B, ou des deux événements A et B simultanément.

Définition 5.7

Soient deux événements A et B. La réalisation de l'événement D, défini par $D = A \cap B$ (lire A **inter** B) entraîne la réalisation de l'événement A et de l'événement B.

Reprenons notre exemple de lancer de dé. À partir de l'univers des possibles $\Omega = \{1,2,3,4,5,6\}$, on peut définir plusieurs types d'événements par union ou intersection d'événements élémentaires ou d'autres événements (► tableau 5.1). Il convient de noter que dans le cas d'un lancer de dé, on ne peut obtenir qu'un seul événement élémentaire à la fois (par exemple 1) par lancer. Dès lors, l'événement $\{\{1\} \cup \{2\}\}$ s'interprète uniquement comme « *on obtient un 1 ou un 2* », car le cas « *on obtient 1 et 2 simultanément* » est impossible. Pour la même raison, l'événement $\{\{1\} \cap \{2\}\}$ est impossible car on ne peut pas obtenir à la fois un 1 et un 2. Pour simplifier les notations, dans ce tableau, nous notons $\{A \cup B\}$ à la place $\{\{A\} \cup \{B\}\}$ lorsque A et B sont des événements élémentaires.

▼ **Tableau 5.1 Exemples d'événements associés à un lancer de dé**

Notation	Interprétation
$A = \{1 \cup 2\}$	On obtient 1 ou 2
$B = \{1 \cap 2\} = \varnothing$	On obtient 1 et 2 : événement impossible
$C = \{1 \cup 2 \cup 3\}$	On obtient 1, 2 ou 3
$D = \{1 \cap \{2 \cup 3\}\} = \varnothing$	On obtient 1 et 2, ou 1 et 3 : événement impossible
$\Omega = \{1 \cup 2 \cup 3 \cup 4 \cup 5 \cup 6\}$	On obtient 1, 2, 3, 4, 5 ou 6 : événement certain

De la combinaison d'événements, nous pouvons déduire les notions d'événements disjoints et d'événements complémentaires.

Définition 5.8

Deux événements A et B sont **disjoints** s'ils n'ont pas d'élément en commun, *i.e.* $A \cap B = \varnothing$. Ces deux événements sont donc incompatibles : la réalisation simultanée de ces événements est impossible.

Définition 5.9

Deux événements A et $\overline{A}$ appartenant à un ensemble B sont **complémentaires** si leur union correspond à B, *i.e.* $A \cup \overline{A} = B$.

La barre horizontale au dessus de la lettre associée à l'événement signifie « *complémentaire de* ». Par exemple, pour un univers Ω = {« bleu », « blanc », « rouge »}, l'événement $\overline{A}$ = {« bleu »} est le complémentaire de l'événement A = {« blanc », « rouge »}.

1.3 Ensemble d'événements

À partir d'événements combinés ou d'événements élémentaires (singletons), il est possible de définir des **ensembles d'événements** ou **parties d'événements**.

Exemple

Pour un univers $\Omega = \{A,B,C\}$, on peut définir les événements $A \cup B$, $B \cap C$ ou $A \cup \{B \cap C\}$. À partir de ces événements et du singleton $\{C\}$, on peut alors construire un ensemble d'événements (ou partie) D tel que :

$$D = \{ \underbrace{\{A \cup B\}}_{\text{événement}}, \underbrace{\{B \cap C\}}_{\text{événement}}, \underbrace{\{A \cup \{B \cap C\}\}}_{\text{événement}}, \underbrace{\{C\}}_{\text{événement}} \} \quad (5.3)$$

Une notion essentielle est celle d'**ensemble de tous les événements réalisables** ou d'ensemble des parties. Cet ensemble recense tous les événements qu'il est possible de définir à partir de l'univers des résultats.

Définition 5.10

L'**ensemble des parties**, noté $\mathcal{P}(\Omega)$, correspond à l'ensemble de tous les événements réalisables à partir des événements élémentaires de l'univers Ω. Par convention $\Omega \in \mathcal{P}(\Omega)$ et $\varnothing \in \mathcal{P}(\Omega)$.

Par convention, l'événement certain (univers) et l'ensemble des événements impossibles appartiennent toujours à l'ensemble des parties $\mathcal{P}(\Omega)$. Attention, il convient de ne pas confondre l'univers de tous les *résultats possibles* Ω et l'ensemble $\mathcal{P}(\Omega)$ de tous les *événements* que l'on peut définir à partir de Ω.

Exemple

Considérons l'exemple d'un lancer de dé à trois faces. L'univers des résultats possibles est $\Omega = \{1,2,3\}$. En effet, le résultat de l'expérience aléatoire, c'est-à-dire du lancer de dé, sera soit 1, soit 2 ou soit 3. En revanche à partir de cet univers de cardinalité égale à 3, on peut construire $2^3 = 8$ événements (ou parties) recensés dans le tableau 5.2.

▼ **Tableau 5.2 Ensemble des événements pour un lancer de dé à trois faces**

Événement	Interprétation	Événement	Interprétation
$\{1\}$	On obtient 1	$\{1 \cup 3\}$	On obtient 1 ou 3
$\{2\}$	On obtient 2	$\{2 \cup 3\}$	On obtient 2 ou 3
$\{3\}$	On obtient 3	$\{1 \cup 2 \cup 3\}$	On obtient 1, 2 ou 3
$\{1 \cup 2\}$	On obtient 1 ou 2	$\varnothing$	Événement impossible

Par convention, on inclut l'ensemble vide dans l'ensemble des parties. Par conséquent, l'ensemble de tous les événements réalisables $\mathcal{P}(\Omega)$ est défini par :

$$\mathcal{P}(\Omega) = \{\{1\},\{2\},\{3\},\{1 \cup 2\},\{1 \cup 3\},\{2 \cup 3\},\{1 \cup 2 \cup 3\},\varnothing\} \quad (5.4)$$

L'événement $\{1 \cup 2 \cup 3\}$, qui s'interprète comme le fait d'obtenir 1, 2 ou 3, peut être noté sous la forme $\{1,2,3\}$ et correspond à l'événement certain (univers) Ω. On peut donc aussi noter l'ensemble des parties sous la forme suivante :

$$\mathcal{P}(\Omega) = \{\{1\},\{2\},\{3\},\{1 \cup 2\},\{1 \cup 3\},\{2 \cup 3\},\Omega,\varnothing\} \quad (5.5)$$

Remarque : Pour un univers des possibles Ω de dimension finie, de cardinalité $card(\Omega) = n$, la cardinalité de l'ensemble des parties $\mathcal{P}(\Omega)$ est égale à :

$$card(\mathcal{P}(\Omega)) = 2^n \quad (5.6)$$

1.4 Tribu d'événements et espace probabilisable

Une **tribu**[2] ou **σ-algèbre** (prononcer sigma-algèbre) sur un univers fini ou infini est un ensemble d'événements de cet univers vérifiant deux principales propriétés : la stabilité par passage au complémentaire et la stabilité par réunion dénombrable.

Définition 5.11

Une **tribu** ou **σ-algèbre** sur l'univers Ω est un sous-ensemble d'événements ou de parties, notée $\mathcal{F}$, vérifiant :

1. $\mathcal{F} \subseteq \mathcal{P}(\Omega)$, $\Omega \in \mathcal{F}$ et $\varnothing \in \mathcal{F}$.
2. L'ensemble $\mathcal{F}$ est stable par **passage au complémentaire** : pour tout événement A de $\mathcal{F}$, l'événement complémentaire $\overline{A}$ appartient à l'ensemble $\mathcal{F}$.
$$\forall A \in \mathcal{F} \text{ alors } \overline{A} \in \mathcal{F} \tag{5.7}$$
3. L'ensemble $\mathcal{A}$ est stable par **réunion dénombrable** : pour toute suite d'événements $(A_n)_{n\in\mathbb{N}}$ appartenant à $\mathcal{F}$, l'union de ces événements appartient à l'ensemble $\mathcal{F}$.
$$(A_n)_{n\in\mathbb{N}} \in \mathcal{F} \text{ alors } \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F} \tag{5.8}$$

Par convention, on note les tribus par des lettres avec une police de caractère dite calligraphique, comme par exemple $\mathcal{A}$, $\mathcal{B}$, $\mathcal{F}$, etc. Le point important de cette définition est que la stabilité par réunion dénombrable garantit que toute union de sous-éléments (événements) de la tribu est équivalente à un autre événement qui appartient lui même à la tribu.

Exemple

Soit un univers $\Omega = \{1,2,3\}$, alors l'ensemble $\mathcal{A} = \{\varnothing,\{1\},\{2,3\},\Omega\}$ est une tribu sur Ω. En effet, cet ensemble appartient à l'ensemble des parties $\mathcal{P}(\Omega)$. De plus, il comprend l'ensemble vide $\varnothing$ et l'événement certain (univers) Ω. Cet ensemble est stable par passage au complémentaire. Par exemple, si l'on pose par exemple $B = \{1\}$, alors $\overline{B} = \{\varnothing,\{2,3\},\Omega\} \in \mathcal{A}$, et il en va de même pour tout sous-ensemble de $\mathcal{A}$. Enfin, si l'on considère par exemple l'union des événements $\{1\}$ et $\{2,3\}$, on obtient un événement $C = \{1 \cup \{2,3\}\} = \Omega \in \mathcal{A}$. On obtient un résultat similaire pour toute union des sous-ensembles de A.

Il existe plusieurs exemples de tribus « évidentes ». La *tribu triviale* (ou tribu grossière) est la plus petite tribu sur Ω. Elle est définie par $\mathcal{F} = \{\varnothing,\Omega\}$. Dans le cas d'un univers fini ou infini dénombrable, une autre tribu « évidente » est donnée par l'ensemble des parties $\mathcal{P}(\Omega)$.

Propriété

Ensemble des parties

Si l'univers Ω est fini ou infini dénombrable, l'ensemble des parties $\mathcal{P}(\Omega)$ est une tribu sur Ω.

2 Le terme de tribu, utilisé en français pour dénommer les σ-algèbres, a été introduit dans un article publié en 1936 par René de Possel.

Les tribus permettent de définir la notion d'**univers probabilisable**[3].

Définition 5.12

Un **univers probabilisable** est un couple $(\Omega,\mathcal{F})$ où $\mathcal{F}$ est une tribu (ou σ-algèbre) sur l'univers Ω.

Dans le cas d'un univers fini ou infini dénombrable, un univers probabilisable est donné par $(\Omega,\mathcal{P}(\Omega))$ puisque $\mathcal{P}(\Omega)$ est une tribu sur Ω.

Exemple

On considère l'expérience aléatoire qui consiste à lancer un dé à trois faces parfaitement équilibrées. L'univers des résultats possibles est $\Omega = \{1,2,3\}$. Comme nous l'avons vu précédemment, l'ensemble des parties $\mathcal{P}(\Omega)$ est défini par :

$$\mathcal{P}(\Omega) = \{\{1\},\{2\},\{3\},\{1\cup 2\},\{1\cup 3\},\{2\cup 3\},\{1\cup 2\cup 3\},\varnothing\} \tag{5.9}$$

Puisque l'univers Ω est fini, l'ensemble des parties $\mathcal{P}(\Omega)$ est une *tribu* (ou σ-algèbre) sur Ω et le couplet $(\Omega,\mathcal{P}(\Omega))$ est un *univers probabilisable*. Un univers probabilisable est un univers de résultats sur lequel nous pouvons définir des *probabilités*.

Remarque : En règle générale (mais pas toujours), dans le cas d'un *univers fini ou infini dénombrable*, on définit les probabilités sur l'univers probabilisable $(\Omega,\mathcal{P}(\Omega))$, où la tribu sur Ω correspond à l'ensemble des parties.

2 Probabilités

2.1 Définition générale d'une probabilité

Une **mesure de probabilité** (ou probabilité) est une application qui associe à tout événement appartenant à une tribu une valeur sur $[0,1]$.

Définition 5.13

Soit $(\Omega,\mathcal{F})$ un univers probabilisable fini ou infini dénombrable. Une **probabilité** (ou mesure de probabilité) est une application $\Pr : \mathcal{F} \to [0,1]$, telle que :

1. $\Pr(\Omega) = 1$.
2. Pour toute suite d'événements disjoints $(A_n)_{n\in\mathbb{N}}$ de $\mathcal{F}$ on a (propriété de **σ-additivité**) :

$$\Pr\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \Pr(A_n) \tag{5.10}$$

Pour tout événement $A \in \mathcal{F}$, le nombre $\Pr(A)$ correspond à la probabilité de l'événement A. Reprenons notre exemple de lancer de dé à trois faces.

3 Les termes d'espace probabilisable, d'espace mesurable ou d'univers mesurable sont aussi souvent employés.

Exemple

On considère l'expérience aléatoire qui consiste à lancer un dé à trois faces parfaitement équilibrées avec $\Omega = \{1,2,3\}$. L'ensemble des parties $\mathcal{P}(\Omega)$ est :

$$\mathcal{P}(\Omega) = \{\{1\}, \{2\}, \{3\}, \{1 \cup 2\}, \{1 \cup 3\}, \{2 \cup 3\}, \{1 \cup 2 \cup 3\}, \varnothing\} \qquad (5.11)$$

Le couplet $(\Omega, \mathcal{P}(\Omega))$ est un *univers probabilisable*, on peut donc lui associer une mesure de probabilité $\Pr : \mathcal{P}(\Omega) \rightarrow [0,1]$, telle que pour tout événement $A \in \mathcal{P}(\Omega)$ il existe une probabilité $\Pr(A) \in [0,1]$. Puisque le dé est parfaitement équilibré, les événements élémentaires $\{1\}$, $\{2\}$ et $\{3\}$ sont équiprobables et leur probabilité est égale à 1/3. On en déduit les probabilités pour tous les événements de $\mathcal{P}(\Omega)$. Le tableau 5.3 synthétise ces $2^3 = 8$ probabilités. On vérifie que la probabilité associée à l'événement certain (univers des résultats Ω) est égale à 1, tandis que la probabilité associée aux événements impossibles $\varnothing$ est égale à 0. On vérifie en outre que cette mesure de probabilité satisfait la propriété de σ-additivité, qui est la conséquence de la propriété de stabilité par réunion de la σ-algèbre $\mathcal{P}(\Omega)$. La probabilité associée à l'union de n'importe quels événements disjoints de la tribu $\mathcal{P}(\Omega)$ est égale à la somme des probabilités des événements. Par exemple :

$$\Pr(\{1\} \cup \{2\}) = \Pr(A_1 \cup A_2) = P(A_1) + P(A_2) = \frac{2}{3} \qquad (5.12)$$

$$\Pr(\{1 \cup 2\} \cup \{3\}) = \Pr(A_4 \cup A_3) = P(A_4) + P(A_3) = 1 \qquad (5.13)$$

▼ Tableau 5.3 **Probabilités pour un lancer de dé à trois faces**

Événement	Probabilité	Événement	Probabilité
$A_1 = \{1\}$	$\Pr(A_1) = 1/3$	$A_5 = \{1 \cup 3\}$	$\Pr(A_5) = 2/3$
$A_2 = \{2\}$	$\Pr(A_2) = 1/3$	$A_6 = \{2 \cup 3\}$	$\Pr(A_6) = 2/3$
$A_3 = \{3\}$	$\Pr(A_3) = 1/3$	$\Omega = \{1 \cup 2 \cup 3\}$	$\Pr(\Omega) = 1$
$A_4 = \{1 \cup 2\}$	$\Pr(A_4) = 2/3$	$\varnothing$	$\Pr(\varnothing) = 0$

Remarque : Si l'univers est *infini non dénombrable*, les probabilités associées aux « événements élémentaires »[4], par exemple $\Pr(\{2\})$, tendent vers 0 puisque la somme infinie de ces probabilités est égale à 1. Seules les probabilités associées à des événements composés du type $\Pr(\{[2,4]\})$, c'est-à-dire la probabilité d'appartenir à l'intervalle de valeurs [2,4], sont non nulles. On dit alors que *la probabilité d'être en point (singleton) est nulle*. C'est ce qui fonde la différence entre les **variables aléatoires continues** et les **variables aléatoires discrètes** (► chapitre 6 sur les variables aléatoires).

La mesure de probabilité nous permet de définir un **univers probabilisé** (ou un espace probabilisé).

Définition 5.14

Un **univers probabilisé** est un triplet $(\Omega, \mathcal{F}, \Pr)$ où $\mathcal{F}$ est une σ-algèbre sur l'univers Ω et $\Pr(.)$ une mesure de probabilité.

4 Au sens strict, le terme d'*événement élémentaire* est inapproprié dans ce cas.

2.2 Définition axiomatique

Dans le cas d'un univers fini, on peut proposer une définition équivalente de la probabilité, dite **définition axiomatique**. Ces **axiomes des probabilités** sont aussi appelés axiomes de Kolmogorov (▶ encadré : les grands auteurs).

Définition 5.15

Soit $(\Omega,\mathcal{F})$ un univers probabilisable fini tel que $\Omega = \{\omega_1,...,\omega_n\}$ et soit $\mathcal{P}(\Omega)$ l'ensemble des parties, avec $\mathcal{F} \subseteq \mathcal{P}(\Omega)$. Une **probabilité** est une application $\text{Pr} : \mathcal{F} \to [0,1]$, telle que :

1. La somme des probabilités associées aux événements élémentaires (ou singletons) ω_i est égale à 1 :

$$\sum_{i=1}^{n} \text{Pr}(\omega_i) = 1 \qquad (5.14)$$

2. La probabilité d'un événement $A \in \mathcal{F}$ est égale à la somme des probabilités associées aux événements élémentaires ω_i qui le constituent :

$$\text{Pr}(A) = \sum_{\omega_i \in A} \text{Pr}(\omega_i) \qquad (5.15)$$

Reprenons notre exemple.

Exemple

On considère l'expérience aléatoire qui consiste à lancer un dé à trois faces parfaitement équilibrées avec $\Omega = \{1,2,3\}$. On vérifie que pour l'*univers probabilisé* $(\Omega,\mathcal{P}(\Omega), \text{Pr})$ décrit précédemment, on a bien :

$$\text{Pr}(\{1\}) + \text{Pr}(\{2\}) + \text{Pr}(\{3\}) = \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1 \qquad (5.16)$$

Par ailleurs la probabilité de tout événement $A \in \mathcal{P}(\Omega)$ est égale à la somme des probabilités associées aux événements élémentaires qui le constituent. Par exemple, pour l'événement $A = \{2 \cup 3\}$, on a :

$$\text{Pr}(\{2 \cup 3\}) = \text{Pr}(\{2\}) + \text{Pr}(\{3\}) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3} \qquad (5.17)$$

De ces définitions, nous pouvons déduire certaines propriétés de la mesure de probabilité.

Propriété

Mesure de probabilité

Soit un univers probabilisé $(\Omega,\mathcal{F}, \text{Pr})$, alors quels que soient les événements A et B appartenant à $\mathcal{F} \subseteq \mathcal{P}(\Omega)$, la mesure de probabilité Pr vérifie :

1. $\text{Pr}(\Omega) = 1$.

2. $\text{Pr}(\varnothing) = 0$.

3. $\text{Pr}\left(\overline{A}\right) = 1 - \text{Pr}(A)$.

4. $\text{Pr}(A \cup B) = \text{Pr}(A) + \text{Pr}(B) - \text{Pr}(A \cap B)$.

5. Si $A \subset B$, alors $\text{Pr}(A) \leq \text{Pr}(B)$.

La première propriété signifie que la probabilité associée à l'événement certain (univers) est égale à 1. La seconde propriété signifie que la probabilité associée à tout événement impossible est nulle. La troisième propriété découle de la propriété de stabilité par passage au complémentaire de la tribu : la probabilité d'un événement est toujours égale à 1 moins la probabilité de son événement complémentaire. La quatrième propriété implique notamment que si deux événements A et B sont incompatibles (ou disjoints), c'est-à-dire si si $A \cap B = \emptyset$, alors :

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) \tag{5.18}$$

Exemple

On considère une famille qui a 2 enfants. Calculons les probabilités associées aux événements A : « deux enfants sont de sexe différent » et B : « il y a au plus une fille ». Dans cette expérience aléatoire, on peut représenter l'univers des possibles par un ensemble de 4 couplets $\{a,b\}$ où a désigne le sexe du premier enfant et b le sexe du deuxième enfant.

$$\Omega = \{\{G,G\}, \{G,F\}, \{F,G\}, \{F,F\}\} \tag{5.19}$$

où G désigne un garçon et F une fille. Si l'on admet que ces événements sont équiprobables, alors les probabilités associées à ces quatre événements élémentaires (singletons) sont égales à 1/4. Calculons la probabilité de l'événement A :

$$\Pr(A) = \Pr(\{F,G\} \cup \{G,F\}) \tag{5.20}$$

$$= \Pr(\{F,G\}) + \Pr(\{G,F\}) - \Pr(\{F,G\} \cap \{G,F\}) \tag{5.21}$$

Puisque les événements élèmentaires $\{F,G\}$ et $\{G,F\}$ sont disjoints (attention à l'ordre des enfants), $\Pr(\{F,G\} \cap \{G,F\}) = 0$. On en déduit que la probabilité de l'événement « deux enfants sont de sexe différent » est égale à :

$$\Pr(A) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \tag{5.22}$$

De la même façon, on peut déterminer la probabilité de l'événement « il y a au plus une fille ». Cet événement signifie qu'il y a soit une fille comme premier enfant, soit une fille comme second enfant, soit aucune fille parmi les deux enfants.

$$\Pr(B) = \Pr(\{F,G\} \cup \{G,F\} \cup \{G,G\}) \tag{5.23}$$

$$= \Pr(\{F,G\}) + \Pr(\{G,F\}) + \Pr(\{G,G\}) \tag{5.24}$$

puisque les trois événements élémentaires sont disjoints deux à deux. Donc :

$$\Pr(B) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4} \tag{5.25}$$

On peut enfin définir le concept de **suite croissante** ou **décroissante** d'événements et la propriété de limite monotone.

Définition 5.16

Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements. On dit que cette suite est **croissante** si $\forall n \in \mathbb{N}, A_n \subset A_{n+1}$ et **décroissante** si $\forall n \in \mathbb{N}, A_{n+1} \subset A_n$.

Pour une suite croissante, on obtient :

$$\Pr\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \Pr(A_n) \tag{5.26}$$

Pour une suite décroissante, on obtient :

$$\Pr\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \Pr(A_n) \tag{5.27}$$

2.3 Fréquence d'événement et probabilité

Nous avons défini une expérience aléatoire comme une expérience pouvant être répétée, en pratique ou en théorie, dans les mêmes conditions. Dans ce contexte, il existe une troisième façon de caractériser (et non de définir) une probabilité associée à un événement en utilisant la **fréquence** d'apparition (ou fréquence empirique) de cet événement. Cette approche est appelée l'**approche fréquentiste**.

Définition 5.17

On considère une expérience aléatoire répétée s fois dans des conditions strictement identiques. La **fréquence** d'apparition de l'événement $A \in \mathcal{P}(\Omega)$ est définie par :

$$F_s(A) = \frac{\text{Nombre de fois où } A \text{ se réalise}}{s} \tag{5.28}$$

Exemple

On considère un lancer de dé à trois faces, avec $\Omega = \{1,2,3\}$, pour lequel on s'intéresse à l'événement « obtenir un 1 ». On suppose que l'on peut répéter le lancer 100 fois exactement dans les mêmes conditions. Si l'on obtient un 1 dans 37 tirages, la fréquence de l'événement $\{1\}$ sera alors égale à $F_{100}(\{1\}) = 37\ \%$.

Dans la vie quotidienne, il est courant de confondre les notions de fréquence et de probabilité. Or ces deux notions ne sont absolument pas équivalentes. Si 80 000 lycéens de terminale sur 100 000 obtiennent leur baccalauréat, cela ne signifie pas que, pour un étudiant donné, la probabilité de réussite est égale à 80 %. La seule relation qui existe entre les deux notions, c'est que la fréquence converge vers la probabilité si l'on peut répéter l'expérience une infinité de fois dans les mêmes conditions.

Propriété

Convergence de la fréquence

Lorsqu'il est possible de réaliser l'expérience aléatoire une infinité de fois dans les mêmes conditions, la **fréquence d'apparition** de tout événement $A \in \mathcal{P}(\Omega)$ converge vers sa probabilité :

$$\lim_{s\to\infty} F_s(A) = \Pr(A) \tag{5.29}$$

Attention, il convient de noter que cette propriété n'est pas une définition de la probabilité. Considérons l'exemple du lancer de pièce avec $\Omega = \{$« pile »,« face »$\}$. Si l'on lance la pièce dans les mêmes conditions un grand nombre de fois et que l'on obtient « pile » dans 48 % des cas, cela veut juste dire que la probabilité de l'événement « pile » peut être approchée par 0,48. La fréquence est donc un moyen de « quantifier » ou d'**estimer** (▶ chapitre 9) la probabilité d'un événement.

Le problème de cette propriété est que l'on ne peut que rarement répéter une expérience aléatoire une infinité de fois dans les mêmes conditions. Par ailleurs, même

si cela est possible en théorie, comment traduire la notion d'infini ? Combien de répétitions de l'expérience sont nécessaires pour obtenir une évaluation précise de la probabilité ? Admettons que 100 000 répétitions de l'expérience soient suffisantes et que l'on obtienne une fréquence de 0,47 pour un événement A. Que se passe-t-il si l'on fait une 100 001e répétition supplémentaire ? Si l'événement A se réalise à nouveau, la fréquence passe alors à 0,47001. Quel est le niveau de la probabilité $\Pr(A)$, 0,47 ou 0,47001 ?

Tout ceci illustre le fait que la fréquence et la probabilité sont des objets de natures très différentes : la fréquence est une **variable aléatoire** tandis que la probabilité est une constante. Nous ne pouvons pas dire que la probabilité correspond ou est définie par la fréquence. C'est seulement dans le cas hypothétique où $s \to \infty$ que les deux objets coïncident : la fréquence **converge** alors vers la probabilité. De façon générale, nous verrons que la fréquence est un **estimateur** de la probabilité (▸ chapitre 9).

3 Probabilité conditionnelle

3.1 Définition de la probabilité conditionnelle

On considère une expérience aléatoire représentée par un univers probabilisé $(\Omega, \mathcal{F}, \Pr)$. On s'intéresse à la probabilité d'un événement $A \in \mathcal{F}$. Mais, alors que l'expérience n'est pas réalisée, c'est-à-dire avant que l'on obtienne le résultat, on obtient une information qui se traduit par un événement $B \in \mathcal{F}$. La probabilité associée à l'événement A doit donc tenir compte de l'événement B. Cette probabilité associée à l'événement A *sachant* l'événement B est appelée **probabilité conditionnelle**.

Définition 5.18

Soit un univers probabilisé $(\Omega, \mathcal{F}, \Pr)$ et soient A et B deux événements appartenant à la tribu $\mathcal{F}$ sur Ω, tels que $\Pr(B) > 0$. La **probabilité conditionnelle** de l'événement A sachant B est définie par :

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \tag{5.30}$$

Le conditionnement dans la mesure de probabilité est indiqué par une barre verticale. À gauche de cette barre figure l'événement pour lequel on cherche à déterminer la probabilité conditionnelle (A) et à droite de celle-ci figure l'événement dit « de conditionnement » (B). On note aussi parfois la probabilité conditionnelle de A sachant B sous la forme $\Pr_B(A)$. Mais quelle que soit la notation adoptée, il convient de remarquer que l'événement de conditionnement a nécessairement une probabilité non nulle pour que la probabilité conditionnelle soit définie. Dit autrement, un événement impossible ne peut pas être un événement de conditionnement.

Exemple

On considère un patient qui peut suivre au choix deux traitements, notés A et B. La probabilité qu'il suive le traitement A est égale à 75 %. On sait par ailleurs que la probabilité de succès du traitement A est égale à 80 %, tandis que celle du traitement B est égale à 90 %. On admet que la probabilité que ce patient soit guéri, événement noté G, est égale à 82,5 %. Déterminons la

probabilité que le patient ait suivi le traitement A sachant qu'il est guéri. D'après la définition de la probabilité conditionnelle, il vient :

$$\Pr(A|G) = \frac{\Pr(A \cap G)}{\Pr(G)} = \frac{\Pr(G|A) \times P(A)}{\Pr(G)} = \frac{0{,}8 \times 0{,}75}{0{,}825} = 0{,}7273 \quad (5.31)$$

La probabilité conditionnelle étant une probabilité, elle vérifie nécessairement la définition générale de la section 2. En particulier, on peut montrer que la probabilité conditionnelle de l'événement certain sachant n'importe quel événement $A \in \mathcal{F}$ est égale à 1.

$$\Pr(\Omega|A) = \frac{\Pr(\Omega \cap A)}{\Pr(A)} = \frac{\Pr(A)}{\Pr(A)} = 1 \quad (5.32)$$

Pour les mêmes raisons, la probabilité conditionnelle d'une union d'événements disjoints $(A_1,\ldots,A_n)$ correspond à la somme des probabilités de ces événements :

$$\Pr\left(\bigcup_{i=1}^{n} A_i | B\right) = \sum_{i=1}^{n} \Pr(A_i|B) \quad (5.33)$$

Par exemple, pour deux événements disjoints $(A_1,A_2) \in \mathcal{F}^2$, on a :

$$\Pr(A_1 \cup A_2|B) = \Pr(A_1|B) + \Pr(A_2|B) \quad (5.34)$$

Compte tenu de la définition de la probabilité conditionnelle, on note immédiatement que pour deux événements de probabilité non nulle A et B, on peut déterminer la probabilité de l'intersection $A \cap B$, dite **probabilité jointe**, de deux façons.

Définition 5.19

Soit un univers probabilisé $(\Omega,\mathcal{F},\Pr)$ et deux événements $(A,B) \in \mathcal{F}^2$, tels que $\Pr(A) > 0$ et $\Pr(B) > 0$. La **probabilité jointe** associée à l'événement $A \cap B$ est définie par :

$$\Pr(A \cap B) = \Pr(A|B) \times \Pr(B) = \Pr(B|A) \times \Pr(A) \quad (5.35)$$

Cette définition peut se généraliser à plus de deux événements par la formule de l'intersection ou formule des **probabilités composées**.

Propriété

Formule de l'intersection

Soit un ensemble d'événements $(A_1,\ldots,A_n) \in \mathcal{F}^n$ tels que $\Pr(A_1 \cap \ldots \cap A_n) > 0$, alors la **probabilité jointe** associée à l'événement $A_1 \cap \ldots \cap A_n$ est définie par :

$$\Pr(A_1 \cap \ldots \cap A_n) = \Pr(A_1) \times \Pr(A_2|A_1) \times \Pr(A_3|A_1 \cap A_2) \times \ldots \times \Pr(A_n|A_1 \cap \ldots \cap A_{n-1}) \quad (5.36)$$

Cette formule des probabilités composées est particulièrement utile pour calculer des probabilités d'intersections, notamment dans le cas d'une *succession d'expériences* aléatoires.

Exemple

On dispose d'une urne contenant $n - 1$ boules noires et 1 boule rouge. On tire au hasard les boules une à une sans remise (il s'agit ici d'une succession d'expériences) et l'on cherche à calculer la probabilité que l'on obtienne la boule rouge à l'issue du $i^{ème}$ tirage avec $1 \leq i \leq n$. On définit A_i comme l'événement « le $i^{ème}$ tirage est un échec » et B_i comme l'événement « le $i^{ème}$ tirage est un succès ». Par définition :

$$\Pr(B_i) = \Pr(A_1 \cap \ldots \cap A_{i-1} \cap B_i) \tag{5.37}$$

Par application de formule de l'intersection, on obtient :

$$\begin{aligned}\Pr(B_i) &= \Pr(A_1) \times \Pr(A_2|A_1) \times \Pr(A_3|A_1 \cap A_2) \times \ldots \\ &\ldots \times \Pr(A_{i-1}|A_1 \cap \ldots \cap A_{i-2}) \times \Pr(B_i|A_1 \cap \ldots \cap A_{i-1})\end{aligned} \tag{5.38}$$

Au premier tirage, il y a n boules dans l'urne et la probabilité d'un échec est égale à :

$$\Pr(A_1) = \frac{n-1}{n} \tag{5.39}$$

Au deuxième tirage, puisqu'il ne reste que $n - 1$ boules au total, la probabilité conditionnelle d'un nouvel échec est égale à :

$$\Pr(A_2|A_1) = \frac{n-2}{n-1} \tag{5.40}$$

Plus généralement, au tirage $i - 1$, il reste $n - i + 2$ boules dans l'urne et la probabilité conditionnelle d'un échec est :

$$\Pr(A_{i-1}|A_1 \cap \ldots \cap A_{i-2}) = \frac{n-i+1}{n-i+2} \tag{5.41}$$

Enfin, au $i^{ème}$ tirage, il ne reste que $n - i + 1$ boules dans l'urne. La probabilité conditionnelle d'obtenir la boule rouge sachant que l'on a échoué jusque-là est égale à :

$$\Pr(B_i|A_1 \cap \ldots \cap A_{i-1}) = \frac{1}{n-i+1} \tag{5.42}$$

Par conséquent, la probabilité de réussite au $i^{ème}$ tirage est égale à $1/n$.

$$\Pr(B_i) = \frac{n-1}{n} \times \frac{n-2}{n-1} \times \ldots \times \frac{n-i+1}{n-i+2} \times \frac{1}{n-i+1} = \frac{1}{n} \tag{5.43}$$

Propriété

Inclusion

Soit un univers probabilisé $(\Omega, \mathcal{F}, \Pr)$ et deux événements $(A, B) \in \mathcal{F}^2$, tels que $A \subset B$, alors :

$$\Pr(B|A) = \Pr(B) - \Pr(A) \tag{5.44}$$

3.2 Système complet et théorème de Bayes

Définition 5.20

Soit $(A_i)_{i \in I}$ une suite finie ou infinie dénombrable d'événements appartenant à la tribu $\mathcal{F}$. On dit que les événements A_i forment un **système complet** si les trois conditions suivantes sont satisfaites :

1. Les événements A_i et A_j sont disjoints, $\forall i \neq j$.
2. $\bigcup_{i \in I} A_i = \Omega$.
3. $\Pr(A_i) > 0, \forall i \in I$.

Dans le cadre d'un système complet d'événements, on peut établir deux résultats fondamentaux qui sont très utilisés dans la pratique :

– la formule des **probabilités totales** ;
– le **théorème de Bayes** ou formule de probabilité des causes.

Propriété

Formule des probabilités totales

Soit $(A_i)_{i \in I}$ un système complet d'événements et soit un événement $B \in \mathcal{F}$, alors :

$$\Pr(B) = \sum_{i \in I} \Pr(B|A_i) \times \Pr(A_i) \tag{5.45}$$

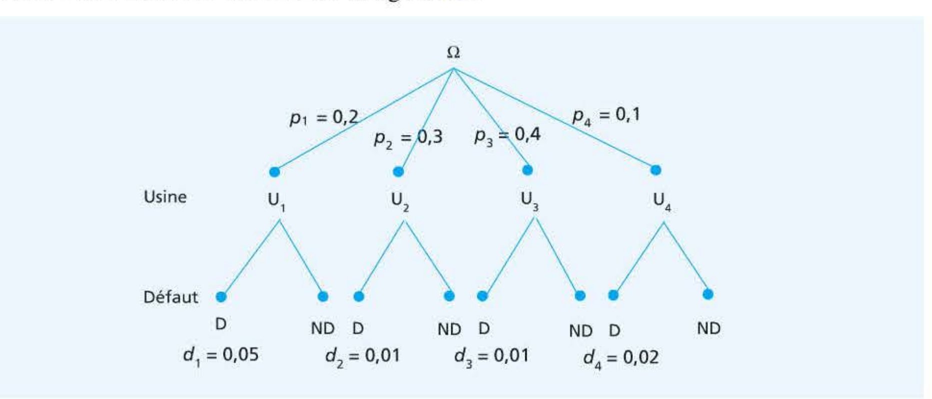
Ce résultat découle de la définition du système complet. En effet, on peut toujours écrire la probabilité $\Pr(B)$ comme $\Pr(B \cap \Omega)$. Dès lors, d'après les propriétés d'un système complet, il vient :

$$\Pr(B) = \Pr(B \cap \Omega) = \Pr\left(B \cap \left(\bigcup_{i \in I} A_i\right)\right) = \Pr\left(\bigcup_{i \in I} (B \cap A_i)\right) \tag{5.46}$$

$$= \sum_{i \in I} \Pr(B \cap A_i) = \sum_{i \in I} \Pr(B|A_i) \times \Pr(A_i) \tag{5.47}$$

Exemple

Une voiture est produite dans quatre usines, notées U_i pour $i = 1,\ldots,4$. On note $p_i = \Pr(U_i)$ la probabilité que la voiture provienne de l'usine U_i, avec $p_1 = 0{,}2$, $p_2 = 0{,}3$, $p_3 = 0{,}4$ et $p_4 = 0{,}1$. Pour chacune de ces usines, la probabilité que la voiture soit défectueuse est notée d_i, avec $d_1 = 0{,}05$, $d_2 = 0{,}01$, $d_3 = 0{,}01$ et $d_4 = 0{,}02$. Bien évidemment, ces probabilités doivent être comprises comme des probabilités conditionnelles de défaut sachant que la voiture est produite dans l'entreprise i et peuvent se noter sous la forme $d_i = \Pr(D|U_i)$ où D représente l'événement « défaut ». Le système complet peut être représenté sous la forme d'un arbre des « défauts » comme sur la figure 5.1.



▲ **Figure 5.1 Arbre des défauts**

Calculons la probabilité qu'une voiture soit défectueuse. D'après la formule des probabilités totales, on a :

$$\Pr(D) = \sum_{i=1}^{4} \Pr(D|U_i) \times \Pr(U_i) \tag{5.48}$$

Ainsi, on montre qu'il y a 1,9 % de chances que la voiture soit défectueuse :

$$\Pr(D) = 0{,}2 \times 0{,}05 + 0{,}3 \times 0{,}01 + 0{,}4 \times 0{,}01 + 0{,}1 \times 0{,}02 = 0{,}019 \tag{5.49}$$

Théorème 5.1

Théorème de Bayes

Soit $(A_i)_{i\in I}$ un système complet d'événements et soit un événement $B \in \mathcal{F}$, tel que $\Pr(B) > 0$, alors $\forall i \in I$:

$$\Pr(A_i|B) = \frac{\Pr(B|A_i) \times \Pr(A_i)}{\sum_{j\in I} \Pr\left(B|A_j\right) \times \Pr\left(A_j\right)} \tag{5.50}$$

Le théorème de Bayes se déduit immédiatement de la formule de la probabilité totale, puisque :

$$\Pr(A_i|B) = \frac{\Pr(A_i \cap B)}{\Pr(B)} = \frac{\Pr(B|A_i) \times \Pr(A_i)}{\sum_{i\in I} \Pr(B|A_i) \times \Pr(A_i)} \tag{5.51}$$

Remarque : Dans le cas particulier où les deux événements A et B forment un système complet, le théorème de Bayes correspond à la définition « intuitive » de la probabilité conditionnelle et peut s'écrire sous la forme :

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \tag{5.52}$$

Le théorème de Bayes est le fondement de la **statistique et de l'économétrie Bayésienne**, dans laquelle la probabilité $\Pr(A_i|B)$ est appelée une probabilité *a posteriori* et $\Pr(A_i)$ est une probabilité *a priori*. La probabilité *a posteriori* sert notamment à mettre à jour ou actualiser les estimations d'une probabilité ou d'un paramètre quelconque, à partir des observations (voir Greene, 2005).

Remarque : Le théorème de Bayes est aussi appelé *formule de probabilité des causes*. En effet, admettons que les événements A_i pour $i \in I$ s'apparentent aux « causes » de l'événement B. La probabilité $\Pr(A_i|B)$ s'interprète alors comme la probabilité que la cause A_i soit responsable de la survenue de l'événement B, sachant que l'événement B s'est réalisé.

Le théorème de Bayes est en effet particulièrement adapté pour identifier les « causes » d'un événement. Reprenons notre exemple de voiture défectueuse.

Exemple

Une voiture est produite dans quatre usines, notées U_i pour $i = 1,\ldots,4$. On note $p_i = \Pr(U_i)$ la probabilité que la voiture provienne de l'usine U_i, avec $p_1 = 0{,}2$, $p_2 = 0{,}3$, $p_3 = 0{,}4$ et $p_4 = 0{,}1$. Pour chacune de ces usines, la probabilité que la voiture soit défectueuse est notée d_i, avec $d_1 = 0{,}05$, $d_2 = 0{,}01$, $d_3 = 0{,}01$ et $d_4 = 0{,}02$. Considérons une voiture défectueuse prise au hasard et calculons la probabilité qu'elle provienne de la $i^{ème}$ usine. D'après le théorème de Bayes, on a :

$$\Pr(U_i|D) = \frac{\Pr(D|U_i) \times \Pr(U_i)}{\sum_{j=1}^{4} \Pr\left(D|U_j\right) \times \Pr\left(U_j\right)} = \frac{\Pr(D|U_i) \times \Pr(U_i)}{\Pr(D)} \quad \forall i = 1,\ldots,4 \tag{5.53}$$

D'après le résultat de l'exercice précédent, nous savons que la probabilité (totale) de défaut est égale à 1,9 %.

$$\Pr(D) = \sum_{j=1}^{4} \Pr\left(D|U_j\right) \times \Pr\left(U_j\right) = 0{,}019 \tag{5.54}$$

Par conséquent, on obtient :

$$\Pr(U_1|D) = \frac{0{,}05 \times 0{,}2}{0{,}019} = 0{,}5263 \qquad \Pr(U_2|D) = \frac{0{,}01 \times 0{,}3}{0{,}019} = 0{,}1579 \tag{5.55}$$

$$\Pr(U_3|D) = \frac{0{,}01 \times 0{,}4}{0{,}019} = 0{,}2105 \qquad \Pr(U_4|D) = \frac{0{,}02 \times 0{,}1}{0{,}019} = 0{,}1053 \tag{5.56}$$

Puisque le système d'événements est complet, la somme de ces probabilités conditionnelles, par construction, est égale à 1. En conclusion, une voiture défectueuse prise au hasard a la plus forte de chance de provenir de l'usine 1.

4 Indépendance

Intuitivement, deux événements A et B sont indépendants lorsque la connaissance de l'un n'apporte aucune information quant à la probabilité de survenue de l'autre, *i.e.* lorsque $\Pr(A|B) = \Pr(A)$ et que $\Pr(B|A) = \Pr(B)$. Ce résultat implique $\Pr(A \cap B) = \Pr(A) \times \Pr(B)$. L'inconvénient de cette définition intuitive est qu'elle n'est valable que pour des événements non impossibles, c'est-à-dire des événements associés à une probabilité strictement positive. Mais lorsqu'un événement est impossible, il est évident qu'il n'a aucun impact sur l'autre. C'est pourquoi, on peut adopter cette définition.

Définition 5.21

Soit un univers probabilisé $(\Omega, \mathcal{F}, \Pr)$ et deux événements $(A, B) \in \mathcal{F}^2$. Les événements A et B sont dit **indépendants** si :

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B) \tag{5.57}$$

Il est important de noter que la définition de l'indépendance est donc relative à une mesure de probabilité Pr. Les événements A et B sont indépendants *pour une certaine mesure* de probabilité. Pour une autre mesure, ils peuvent ne pas être indépendants. Par ailleurs, deux événements peuvent être indépendants dans une expérience aléatoire, et non indépendants dans une autre.

Exemple

On considère une famille qui a 2 enfants. Dans ce cas, les événements A : « deux enfants sont de sexe différent » et B : « il y a au plus une fille » ne sont pas indépendants. En effet, l'univers des possibles est défini par :

$$\Omega = \{\{G,G\}, \{G,F\}, \{F,G\}, \{F,F\}\} \tag{5.58}$$

où G désigne un garçon et F une fille. Si ces quatre événements élémentaires sont équiprobables, alors :

$$\Pr(A) = \Pr(\{F,G\} \cup \{G,F\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \tag{5.59}$$

$$\Pr(B) = \Pr(\{F,G\} \cup \{G,F\} \cup \{G,G\}) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4} \tag{5.60}$$

$$\Pr(A \cap B) = \Pr(\{F,G\} \cup \{G,F\}) = \frac{1}{2} \tag{5.61}$$

Par conséquent, on vérifie que les événements A et B ne sont pas indépendants puisque :

$$\Pr(A \cap B) = \frac{1}{2} \neq \Pr(A) \times \Pr(B) = \frac{3}{8} \tag{5.62}$$

En revanche, ces deux événements sont indépendants lorsque l'on considère une famille avec 3 enfants (nouvelle expérience aléatoire). Dans ce cas, l'univers des possibles comprend $2^3 = 8$ cas possibles. Chaque événement élémentaire a une probabilité égale à 1/8.

$$\Pr(A) = 1 - \Pr(\{G,G,G\} \cup \{F,F,F\}) = 1 - \frac{1}{8} - \frac{1}{8} = \frac{3}{4} \tag{5.63}$$

$$\Pr(B) = \Pr(\{F,G,G\} \cup \{G,F,G\} \cup \{G,G,F\} \cup \{G,G,G\}) \tag{5.64}$$

$$= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2} \tag{5.65}$$

La probabilité jointe est égale à :

$$\Pr(A \cap B) = \Pr(\{F,G,G\} \cup \{G,F,G\} \cup \{G,G,F\}) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8} \tag{5.66}$$

Dans cette expérience, on vérifie que les événements A et B sont indépendants puisque :

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B) = \frac{3}{4} \times \frac{1}{2} = \frac{3}{8} \tag{5.67}$$

Mais dans la plupart des cas, nous n'avons pas à démontrer que deux événements sont indépendants. La configuration de l'expérience aléatoire nous permet généralement de *postuler* l'indépendance, typiquement lorsqu'il y aucun lien de causalité entre les réalisations.

Exemple

On considère une expérience aléatoire consistant à lancer un dé à 6 faces *deux fois de suite*. L'univers des possibles est $\Omega = \{1,2,3,4,5,6\}$. Il est clair que les événements « obtenir un 1 au premier lancer », noté A, et « obtenir un 6 au deuxième », noté B, sont indépendants par construction.

$$\Pr(A) = \frac{1}{6} \qquad \Pr(B) = \frac{1}{6} \tag{5.68}$$

La probabilité d'obtenir un 1 au premier lancer et un 6 au deuxième est :

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} \tag{5.69}$$

Plus généralement, on peut définir l'indépendance mutuelle de n événements de la façon suivante.

Définition 5.22

Soit un univers probabilisé $(\Omega,\mathcal{F},\Pr)$ et n événements $(A_1,\ldots,A_n) \in \mathcal{F}^n$. Les événements $A_1,\ldots,A_n$ sont dits **mutuellement indépendants** si :

$$\Pr\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \Pr(A_i) \tag{5.70}$$

Si les $A_1,\ldots,A_n$ sont mutuellement indépendants, tout événement A_i est indépendant des événements A_j pour $j \neq i$ et de toute union ou intersection de ces événements.



2 questions à

Damien **Deballon**

Consultant ERS (Enterprise Risk Service) chez Deloitte



Quel est votre parcours professionnel et votre mission actuelle au sein du cabinet Deloitte ?

À l'issue de mon master d'Économétrie et de Statistique appliquée obtenu à l'Université d'Orléans, j'ai été embauché en 2012 en tant que consultant chez Deloitte, l'un des quatre grands cabinets internationaux d'audit et de conseil. J'ai intégré la branche ERS (Enterprise Risk Services) et le service IT Advisory IS/SP (Industries et Services). Une partie de mon activité est consacrée à l'audit en support des missions CAC (commissaire aux comptes), notamment aux problématiques de détection de fraudes et d'anomalies sur le cycle comptable, le cycle achat/vente et le cycle paie. Une autre partie de mon activité s'inscrit dans le cadre du développement de la partie Data&Analytics au sein de Deloitte avec notamment des missions de conseil liées aux problématiques Big Data, à la qualité et la gouvernance des données.

En quoi les notions statistiques de probabilité et de variable aléatoire sont-elles fondamentales dans le cadre de votre activité ?

Du fait de leur caractère aléatoire, les typologies de fraudes peuvent être difficilement identifiables pour une entreprise. Ainsi, dans le cadre de nos analyses de fraudes, nous nous appuyons sur des notions de probabilités de survenance d'événements. Les lois de probabilité apparaissent donc le plus souvent comme le résultat d'un processus de modélisation ajouté à la création d'indicateurs spécifiques pour qualifier la fraude. De plus, la notion de significativité des résultats est fondamentale pour tenir compte du fait que la modélisation du processus de fraude reste complexe et que de nombreux cas, de faux positifs ou faux négatifs peuvent nuire à l'interprétation des cas identifiés. ■

Les points **clés**

- Une expérience aléatoire est une expérience qui peut être répétée dans les mêmes conditions et qui ne donne pas forcément le même résultat à chaque répétition.

- L'univers des possibles correspond à l'ensemble de tous les résultats qui peuvent être obtenus au cours d'une expérience aléatoire.

- L'ensemble des parties correspond à l'ensemble de tous les événements réalisables associés à un univers des possibles.

- Un univers probabilisé est un triplet déterminé par un univers des possibles, une tribu et une mesure de probabilité.

- La formule des probabilités totales permet d'exprimer la probabilité d'un événement en fonction des probabilités conditionnelles des événements d'un système complet.

- Le théorème de Bayes ou formule de probabilité des causes permet de caractériser la probabilité conditionnelle d'un événement dans un système complet.

- Deux événements sont indépendants si leur probabilité jointe est égale au produit de leurs probabilités marginales.

ÉVALUATION

► Corrigés sur www.dunod.com

QCM

Pour chacune des questions suivantes, indiquer si les affirmations sont vraies ou fausses (il peut y avoir plusieurs réponses vraies pour chaque question).

1 Expérience aléatoire

a. Une expérience aléatoire est une expérience qui peut être répétée plusieurs fois dans les mêmes conditions.

b. Le résultat d'une expérience aléatoire est un événement élémentaire.

c. L'univers des possibles d'une expérience aléatoire est l'ensemble des résultats possibles.

d. L'univers des possibles se compose d'événements élémentaires.

e. L'univers des possibles est infini.

2 Événement

a. Un événement peut toujours s'exprimer sous la forme d'une combinaison (union ou intersection) de singletons.

b. Un événement certain correspond à l'univers des possibles.

c. Deux événements disjoints ont un élément en commun.

d. L'union d'un événement et son complémentaire correspond à l'univers des possibles.

e. Un ensemble d'événements comprend toujours l'ensemble vide.

3 Ensemble des parties et tribu

a. L'ensemble des parties correspond à l'univers des possibles.

b. L'ensemble des parties comprend toujours l'ensemble vide et l'univers des possibles.

c. Une tribu appartient nécessairement à l'ensemble des parties.

d. Si un événement appartient à une tribu, son complémentaire appartient aussi à cette tribu.

e. Si une suite d'événements appartient à une tribu, l'intersection de ces événements appartient elle aussi à cette tribu.

4 Mesure de probabilité

a. Une mesure de probabilité peut être définie sur un univers probabilisé.

b. Une mesure de probabilité est une application d'une tribu vers le segment $[0,1]$.

c. Une mesure de probabilité peut être appliquée à l'ensemble des parties.

d. Dans le cas d'un univers fini, la somme des probabilités associées aux événements élémentaires est toujours égale à 1.

e. Soient deux événements A et B tels que $B \subset A$, alors $\Pr(A) \leq \Pr(B)$.

Exercices

5 Probabilité et dénombrement

Lors d'un examen, le professeur propose un QCM composé de 4 questions. Pour chaque question, il y a 5 réponses et l'étudiant doit choisir l'une d'entre elles. On admet que pour chaque question, une seule réponse est exacte.

1. Déterminer le nombre de grilles-réponses possibles.

2. Quelle est la probabilité qu'un étudiant réponde au hasard correctement à *au moins* 2 questions et obtienne ainsi *au moins* la moyenne sur cet exercice ?

6 Probabilité et dénombrement

Lors d'une loterie, 50 billets sont vendus. Seulement 2 billets sont gagnants. Si l'on achète 4 billets, quelle est la probabilité de gagner au moins un lot ?

7 Suite d'événements et probabilité totale

Un fumeur décide d'arrêter. On suppose que si cette personne n'a pas fumé le jour n, alors la probabilité qu'elle fume le jour suivant est égale à 0,1. Mais si cette personne fume le jour n, sa probabilité de fumer le jour suivant est égale à 0,8.

1. Exprimer la probabilité que cette personne fume le jour $n+1$ en fonction de la probabilité qu'elle fume le jour n.
2. Déterminer la limite de cette probabilité avec n. Est-ce que cette personne va s'arrêter de fumer ?

8 Probabilité et indépendance

Une entreprise vend deux produits A et B. Sur sa zone de chalandise (population), la probabilité d'achat du produit A est égale à p_A et la probabilité d'achat du produit B est égale à p_B. On suppose que les décisions d'achat des deux produits sont indépendantes.

1. Pour un individu de la population, quelle est la probabilité d'achat des deux produits ?
2. Pour un individu de la population, quelle est la probabilité d'achat de l'un ou de l'autre produit ?

9 Probabilité et dénombrement

Une urne contient $N = 100$ boules dont $N_B = 75$ boules blanches et $N_R = 25$ boules rouges. On fait $n = 50$ tirages avec remise dans l'urne.

1. Soit l'événement E_k « on tire k boules rouges » avec $0 \leq k \leq n$. De façon générale, montrez que la probabilité $\Pr(E_k)$ est égale à :

$$\Pr(E_k) = C_n^k \left(\frac{N_R}{N}\right)^k \left(1 - \frac{N_R}{N}\right)^{n-k} \tag{5.71}$$

2. Montrez que la probabilité de tirer $k = 10$ boules rouges est égale à 9,85 %.

10 Probabilité conditionnelle

Une entreprise reçoit un lot de pièces détachées qui peut comporter un certain nombre de pièces défectueuses. En présence de pièces défectueuses, le lot est dit défectueux et il est rejeté. On admet que la probabilité qu'une pièce soit défectueuse est égale à 5 %. Afin de décider si l'on doit ou non accepter un lot, l'entreprise met en place une procédure de détection. Les résultats de cette procédure montrent que si le lot est défectueux, le test conduit au rejet du lot avec une probabilité de 98 %. Lorsque le lot est effectivement non défectueux, le test conduit (à tort) au rejet du lot avec une probabilité de 4 %.

1. Quelle est la probabilité qu'un lot soit effectivement défectueux si le test conduit au rejet du lot ?
2. Quelle est la probabilité qu'un lot soit valide si le test conduit au rejet du lot ?
3. Quelle est la probabilité qu'un lot soit défectueux si le test ne conduit pas au rejet du lot ?
4. Quelle est la probabilité qu'un lot soit valide si le test ne conduit pas au rejet du lot ?

11 Probabilité conditionnelle

Dans une classe, on distingue deux types d'étudiants suivant leur filière d'origine. Les étudiants ayant suivi la filière A ont une probabilité de 30 % d'obtenir une mention bien à leur examen, tandis que ceux issus de la filière B ont une probabilité de 20 %. La probabilité qu'un étudiant pris au hasard soit issu de la filière A est égale à 70 %. Quelle est la probabilité qu'un étudiant ayant obtenu une mention bien soit issu de la formation A ?

12 Probabilité conditionnelle

On considère un individu qui se rend régulièrement au cinéma. Soit A_i l'événement « l'individu se rend au cinéma le jour i » avec $\Pr(A_1) = p_1$ donné. On suppose que si un jour cet individu se rend au cinéma, le jour suivant il a une probabilité de 1/8 de s'y rendre aussi. Si l'individu ne se rend pas au cinéma le jour i, il y a une probabilité de 3/8 qu'il s'y rende le jour suivant.

1. Exprimer $\Pr(A_{n+1})$ en fonction de $\Pr(A_n)$.
2. Quelle est la probabilité de l'événement « l'individu se rend au cinéma le jour i » ?

Chapitre 6

La mise en œuvre de la notion de probabilité dans la vie économique passe par l'utilisation de **variables aléatoires**. Par exemple, si une banque cherche à modéliser le risque de défaut associé à un prêt consenti à l'un de ses clients, elle représente généralement ce risque par une variable binaire prenant la modalité 1 si le client connaît un défaut de paiement et 0 sinon. Bien évidemment, cette variable est une variable aléatoire puisque le défaut d'un client n'est pas connu à l'avance. Ainsi, une variable aléatoire n'est rien d'autre qu'une **application mesurable**, c'est-à-dire une sorte de « fonction », d'un univers probabilisé vers un univers des **réalisations** probabilisables. Aux événements issus de l'expérience aléatoire (défaut ou non défaut), on associe des réalisations (1 ou 0) de la variable aléatoire.

Le fait que cette application soit mesurable implique que l'on peut affecter des probabilités à toutes les réalisations et donc qu'il est possible de caractériser la **loi de probabilité** (ou distribution) de la variable aléatoire. Par exemple, la variable « défaut » peut être associée à une loi de Bernoulli. On distingue les variables aléatoires discrètes des variables aléatoires continues. Mais, quel que soit le type de variable considéré, la loi de probabilité peut être toujours représentée de trois façons équivalentes :

(i) par sa **fonction de densité** ou sa **fonction de masse** suivant les cas ;

(ii) par sa **fonction de répartition** ;

(iii) par la population de ses **moments**.

LES GRANDS AUTEURS



Carl Friedrich Gauss (1777-1855)



Carl Friedrich Gauss est un mathématicien allemand qui fut à l'origine de contributions majeures non seulement en mathématiques, mais aussi en astronomie et en physique. Dans le domaine des statistiques, il établit les bases de ce que l'on appellera plus tard la loi normale, dite loi de Gauss ou de Laplace-Gauss.

Dans un ouvrage publié en 1809 et consacré au mouvement des corps célestes, il introduisit la procédure d'estimation des moindres carrés (▸ chapitre 2) qui est aujourd'hui sans conteste la méthode d'estimation la plus utilisée. Afin de prouver les bonnes propriétés de cette méthode d'estimation, il dut supposer que les erreurs de mesure étaient distribuées selon une loi particulière, continue, symétrique et d'espérance nulle. C'est ainsi qu'apparut la fameuse loi normale... ■

Variable aléatoire

Plan

Pré-requis

- **Connaître** la notion de probabilité (▸ chapitre 5).
- **Connaître** les bases du calcul intégral.
- **Connaître** les notions de base du calcul matriciel.

Objectifs

- **Définir** la notion de variable aléatoire discrète ou continue.
- **Introduire** la notion de loi de probabilité.
- **Connaître** la signification d'une fonction de densité ou d'une fonction de masse.
- **Savoir** utiliser une fonction de répartition et un quantile.
- **Définir** la notion de moments.
- **Introduire** les notions de loi conditionnelle et de loi jointe.
- **Définir** la notion d'indépendance statistique.

1 Définition générale

On considère une expérience aléatoire et l'on désigne par Ω l'univers des résultats possibles (► chapitre 5).

Définition 6.1

Soient $(\Omega, \mathcal{F}, \Pr)$ un univers probabilisé et $(X(\Omega), \mathcal{E})$ un univers probabilisable. On appelle **variable aléatoire** toute *application mesurable*, notée X, de Ω vers $X(\Omega)$:

$$\forall x \in \mathcal{F} \qquad X^{-1}(x) \in \mathcal{E} \tag{6.1}$$

Cette définition théorique peut sembler aride, mais elle est en fait très simple à comprendre. Une variable aléatoire est une **application**[1], c'est-à-dire une sorte de « fonction », qui pour chaque événement de l'univers des possibles Ω associe une valeur appartenant à un univers $X(\Omega)$. Plus généralement, l'application X associe à tout événement de la tribu $\mathcal{F}$ sur Ω, une valeur appartenant à la tribu $\mathcal{E}$ sur $X(\Omega)$. Cette valeur peut être *numérique* (on parle alors de variable aléatoire **quantitative**) ou *non numérique* (on parle alors de variable aléatoire **qualitative**).

Définition 6.2

On appelle **réalisations**, les valeurs prises par la variable aléatoire X. L'univers $X(\Omega)$ correspond à l'univers des réalisations.

Exemple

On considère l'expérience aléatoire qui consiste à lancer un dé à 6 faces. L'univers des résultats possibles est alors défini par $\Omega = \{1,2,3,4,5,6\}$. Définissons une variable aléatoire, notée X, comme une application qui prend la réalisation 10 lorsque le résultat du lancer est un nombre pair et 20 lorsque le résultat est un nombre impair. La variable aléatoire X est dite quantitative. Si la variable aléatoire X prend les réalisations « pair » ou « impair », cette variable est dite *qualitative*. Dans les deux cas, on a défini une application associant à tout événement de Ω, un élément de l'univers des réalisations $X(\Omega) = \{10,20\}$ dans le cas de la variable aléatoire quantitative et $X(\Omega) = \{$« pair »,« impair »$\}$ dans le cas de la variable aléatoire qualitative.

Ainsi, la variable aléatoire X est une sorte de fonction qui « *transforme* » les résultats d'une expérience aléatoire définis sur Ω en des réalisations définies sur $X(\Omega)$. Mais le point essentiel c'est que l'on est capable de déterminer les probabilités associées à ces réalisations. En effet, si $(\Omega, \mathcal{F}, \Pr)$ est un univers probabilisé cela signifie que l'on peut attribuer une probabilité à tout événement de la tribu $\mathcal{F}$. Dit autrement, on est capable d'attribuer une probabilité non seulement aux événements élémentaires de Ω, mais aussi à tous les événements combinés (union, intersection, etc.) appartenant à la tribu $\mathcal{F}$ (► chapitre 5). Par conséquent, si pour chaque événement de Ω, on associe *via* la variable aléatoire X, une réalisation de cette variable dans $X(\Omega)$, il est aussi possible

1 De façon générale en mathématiques, une application est une relation entre deux ensembles pour laquelle chaque élément du premier (ensemble de départ) est relié à un unique élément du second (ensemble d'arrivée). Cette notion est légèrement différente de la notion de fonction.

de calculer les probabilités associées à ces réalisations. Plus précisément, il est non seulement possible de déterminer une probabilité pour chaque réalisation élémentaire de X appartenant à l'univers $X(\Omega)$, mais aussi pour toutes les combinaisons possibles de ces réalisations (union, intersection, etc.) appartenant à la tribu $\mathcal{E}$. On peut donc définir une **mesure de probabilité** sur l'univers probabilisable $(X(\Omega), \mathcal{E})$.

Exemple

On considère l'expérience aléatoire qui consiste à lancer un dé à 3 faces. L'univers des résultats possibles est alors défini par $\Omega = \{1,2,3\}$. Une tribu sur Ω est donnée par l'ensemble des parties $\mathcal{P}(\Omega)$:

$$\mathcal{F} = \mathcal{P}(\Omega) = \{\{1\}, \{2\}, \{3\}, \{1 \cup 2\}, \{1 \cup 3\}, \{2 \cup 3\}, \{1 \cup 2 \cup 3\}, \varnothing\} \qquad (6.2)$$

Définissons une variable aléatoire X comme une application qui prend la valeur « pair » lorsque le résultat du lancer est un nombre pair et « impair » dans le cas contraire. L'univers des réalisations de X est défini par $X(\Omega) = \{\text{« pair »}, \text{« impair »}\}$. Une tribu sur $X(\Omega)$ est donnée par l'ensemble des parties $\mathcal{P}(X(\Omega))$:

$$\mathcal{E} = \mathcal{P}(X(\Omega)) = \{\{\text{« pair »}\}, \{\text{« impair »}\}, \{\text{« pair »} \cup \text{« impair »}\}, \varnothing\} \qquad (6.3)$$

Si le dé est équilibré, tous les événements de l'univers Ω sont équiprobables, $\Pr(\{1\}) = \Pr(\{2\}) = \Pr(\{3\}) = 1/3$ et l'on connait les probabilités associées à tous les événements de la tribu $\mathcal{F}$ sur Ω. Ces probabilités sont représentées dans le tableau 6.1.

▼ Tableau 6.1 Probabilités pour un lancer de dé à trois faces

Événement	Probabilité	Événement	Probabilité
$A_1 = \{1\}$	$\Pr(A_1) = 1/3$	$A_5 = \{1 \cup 3\}$	$\Pr(A_5) = 2/3$
$A_2 = \{2\}$	$\Pr(A_2) = 1/3$	$A_6 = \{2 \cup 3\}$	$\Pr(A_6) = 2/3$
$A_3 = \{3\}$	$\Pr(A_3) = 1/3$	$\Omega = \{1 \cup 2 \cup 3\}$	$\Pr(\Omega) = 1$
$A_4 = \{1 \cup 2\}$	$\Pr(A_4) = 2/3$	$\varnothing$	$\Pr(\varnothing) = 0$

On peut alors en déduire une probabilité pour tous les événements de la tribu $\mathcal{E}$ sur l'univers $X(\Omega)$ des réalisations de la variable aléatoire X (► tableau 6.2).

▼ Tableau 6.2 Probabilités associées aux réalisations

Réalisation de X	Probabilité
$E_1 = \{\text{« pair »}\}$	$\Pr(E_1) = \Pr(A_2) = 1/3$
$E_2 = \{\text{« impair »}\}$	$\Pr(E_2) = \Pr(A_1 \cup A_3) = 2/3$
$E_3 = \{\{\text{« pair »}\} \cup \{\text{« impair »}\}\}$	$\Pr(E_3) = 1$
$\varnothing$	$\Pr(\varnothing) = 0$

Mais pour que cette intuition soit valide, il faut que la variable aléatoire soit définie comme une **application mesurable.** Qu'est-ce qu'une application mesurable ?

Définition 6.3

Considérons deux univers A et B munis de leurs tribus respectives $\mathcal{A}$ et $\mathcal{B}$. On dit que l'**application** f de A vers B est **mesurable** si l'image réciproque par f de tout événement de la tribu $\mathcal{B}$ (tribu de l'univers d'arrivée) est incluse dans la tribu $\mathcal{A}$ (tribu de l'univers de départ) :

$$\forall b \in \mathcal{B} \qquad f^{-1}(b) \in \mathcal{A} \tag{6.4}$$

Ainsi une variable aléatoire X est une application mesurable si (*i*) pour chaque événement de la tribu $\mathcal{E}$ définie sur l'univers des réalisations $X(\Omega)$ on peut, en inversant le sens de l'application X, « remonter » à un événement sur Ω, et (*ii*) cet événement appartient à la tribu $\mathcal{F}$ définie sur Ω. C'est le sens de l'équation (6.1) de la définition générale d'une variable aléatoire. On peut vérifier que c'est le cas dans notre exemple de lancer de dé à trois faces (▸ tableaux 6.1 et 6.2). Pourquoi est-ce si important de « remonter » à un événement qui appartienne à la tribu $\mathcal{F}$ sur Ω ? Tout simplement parce que ces événements sont probabilisés. Si à toute réalisation ou combinaison de réalisations correspond un événement sur $\mathcal{F}$, on peut lui associer une probabilité. Par conséquent, si X est une application mesurable, il est possible d'affecter une probabilité à toutes les réalisations de $X(\Omega)$ et à toutes les combinaisons de ces réalisations appartenant à $\mathcal{E}$.

Ces probabilités définissent la **loi de probabilité** (ou distribution ou loi) de la variable X. On distingue deux types de variables aléatoires suivant que l'univers des réalisations $X(\Omega)$ est dénombrable (fini ou infini) ou non dénombrable (infini) :

– les variables aléatoires **discrètes** ;

– les variables aléatoires **continues**.

2 Variables aléatoires discrètes

Une **variable aléatoire discrète** est une variable aléatoire qui peut prendre des réalisations discrètes, c'est-à-dire non continues. Plus formellement, la définition[2] d'une variable aléatoire discrète est la suivante.

Définition 6.4

Soit $(\Omega, \mathcal{F}, \Pr)$ un univers probabilisé fini ou infini dénombrable. On appelle **variable aléatoire discrète** X toute application mesurable $X : \Omega \rightarrow X(\Omega)$ telle que $\forall x_i \in X(\Omega)$:

$$\Pr(X = x_i) = \Pr(\{\omega \in \Omega \; ; \; X(\omega) = x_i\}) \tag{6.5}$$

Le terme $\Pr(X = x_i)$ se lit comme « la probabilité que la variable aléatoire X prenne la réalisation x_i ».

Remarque : Par convention, on note la variable aléatoire avec une lettre majuscule (par exemple X) et sa réalisation avec une lettre minuscule (par exemple x).

2 On trouve aussi parfois la définition équivalente $\forall x_i \in X(\Omega)$, $X^{-1}(x_i) = \{\omega \in \Omega \; ; \; X(\omega) = x_i\} \in \mathcal{F}$, qui correspond à la définition d'une application mesurable.

Cette définition indique tout simplement que la probabilité que la variable X prenne la valeur x_i correspond à la probabilité de l'union de tous les événements ω de l'univers des résultats Ω qui correspondent à la valeur x_i dans l'univers des réalisations $X(\Omega)$. Dit autrement, l'application mesurable X permet de déterminer les probabilités associées aux réalisations $x_i \in X(\Omega)$. Notons qu'en général la tribu $\mathcal{F}$ sur Ω correspond à l'ensemble des parties $\mathcal{P}(\Omega)$.

Remarque : Dans la suite de ce chapitre, nous supposerons que la variable aléatoire est toujours *quantitative*, c'est-à-dire que ses réalisations $x_1, \ldots, x_n, \ldots$ sont des nombres. En effet, dans le cas d'une variable qualitative (par exemple qui prend des réalisations du type « bleu », « blanc », « rouge »), il est toujours possible par un **codage** de se ramener à une variable quantitative (par exemple en posant 1 pour « bleu », 2 pour « blanc » et 3 pour « rouge »). De plus, pour simplifier les notations, nous supposerons que les réalisations x_i sont **ordonnées** suivant l'indice i :

$$x_1 < x_2 < \ldots < x_n < x_{n+1} < \ldots \tag{6.6}$$

2.1 Loi de probabilité

On caractérise une variable aléatoire discrète par sa **loi de probabilité** (ou loi de distribution).

Définition 6.5

L'application $\Pr(X = x_i)$ définie pour toutes les réalisations $x_i \in X(\Omega)$ s'appelle la **loi de probabilité** de la variable discrète X. Puisque les réalisations x_i forment un système complet :

$$\sum_{x_i \in X(\Omega)} \Pr(X = x_i) = 1 \tag{6.7}$$

La loi de probabilité d'une variable discrète n'est rien d'autre que la « liste » des probabilités $\Pr(X = x_1)$, $\Pr(X = x_2), \ldots, \Pr(X = x_n), \ldots$, associées à tous les évènements de l'univers des réalisations $X(\Omega)$. Par définition, la somme de ces probabilités pour toutes les réalisations de l'univers $X(\Omega)$ est toujours égale à 1. Si le nombre de réalisations, noté n, appartenant à l'univers $X(\Omega) = \{x_1, \ldots, x_n\}$ est fini, l'équation (6.7) peut se réécrire sous la forme :

$$\sum_{i=1}^{n} \Pr(X = x_i) = 1 \tag{6.8}$$

Si en revanche, l'univers des réalisations $X(\Omega)$ est infini dénombrable, l'équation (6.7) devient :

$$\lim_{n \to \infty} \sum_{i=1}^{n} \Pr(X = x_i) = 1 \tag{6.9}$$

Remarque : L'univers des réalisations $X(\Omega)$ est aussi appelé **support** de la loi de probabilité de X.

La loi de probabilité de la variable aléatoire X peut être caractérisée par la liste des probabilités $\Pr(X = x_i)$ pour *toutes* les réalisations $x_i \in X(\Omega)$. Bien évidemment lorsque le support de la loi de X est de grande dimension (finie) ou de dimension infinie, une telle représentation n'est plus possible en pratique. On utilise alors la **fonction de masse** associée à la loi de probabilité.

Définition 6.6

La **fonction de masse** (*probability mass function* ou pmf en anglais) est la fonction, notée $f_X(x_i)$, qui à toute réalisation $x_i \in X(\Omega)$ associe la probabilité $\Pr(X = x_i)$:

$$f_X(x_i) = \Pr(X = x_i) \quad \forall x_i \in X(\Omega) \tag{6.10}$$

Par convention, on note la fonction de masse par une lettre en minuscule et l'on indique le « nom » de la variable aléatoire en indice avec une lettre majuscule. Considérons deux exemples de lois de probabilité. La première correspond à une variable aléatoire admettant un univers fini de réalisations, tandis que dans le cas de la seconde, cet univers est infini mais dénombrable.

Exemple

On considère une variable aléatoire discrète Y ayant pour support $Y(\Omega) = \{0,1,\ldots,n\}$ et distribuée selon une **loi binomiale** (▶ focus : les lois usuelles), notée $\mathcal{B}(n,p)$. Sa fonction de masse est donnée par la formule suivante :

$$f_Y(y_i) = \Pr(Y = y_i) = C_n^{y_i} \times p^{y_i} \times (1-p)^{n-y_i} \quad \forall y_i \in Y(\Omega) \tag{6.11}$$

où p est un paramètre tel que $p \in [0,1]$ et $C_n^{y_i}$ est la combinaison de y_i parmi n :

$$C_n^{y_i} = \binom{n}{y_i} = \frac{n!}{y_i! \times (n-y_i)!} \tag{6.12}$$

où ! désigne la factorielle. Si l'on pose par exemple $p = 0{,}5$ et $n = 2$, l'univers des réalisations devient $Y(\Omega) = \{0,1,2\}$ et la fonction de masse s'écrit :

$$f_Y(y_i) = \Pr(Y = y_i) = \frac{2!}{y_i! \times (2-y_i)!} \times 0{,}5^{y_i} \times 0{,}5^{2-y_i} \quad \forall y_i \in \{0,1,2\} \tag{6.13}$$

Les probabilités associées sont égales à :

$$f_Y(0) = \Pr(Y = 0) = \frac{2!}{0! \times 2!} \times 0{,}5^0 \times 0{,}5^2 = 0{,}25 \tag{6.14}$$

$$f_Y(1) = \Pr(Y = 1) = \frac{2!}{1! \times 1!} \times 0{,}5^1 \times 0{,}5^1 = 0{,}50 \tag{6.15}$$

$$f_Y(2) = \Pr(Y = 2) = \frac{2!}{2! \times 0!} \times 0{,}5^2 \times 0{,}5^0 = 0{,}25 \tag{6.16}$$

On vérifie que $\sum_{i=0}^{2} \Pr(Y = i) = 1$. La loi de la variable aléatoire peut être représentée soit par le triplet de probabilités $\Pr(Y = 0)$, $\Pr(Y = 1)$ et $\Pr(Y = 2)$, soit par l'équation (6.13) de la fonction de masse. Les deux représentations sont équivalentes.

Exemple

On considère une variable aléatoire discrète Z ayant pour univers des réalisations $Y(Z) = \mathbb{N}$, l'ensemble des entiers naturels $\{0,1,2,\ldots\}$. Cet univers est de taille infinie, mais il est dénombrable. On admet que la variable Z suit une de **loi de Poisson** de paramètre $\lambda > 0$ (▶ focus : les lois usuelles) telle que :

$$f_Z(z) = \Pr(Z = z) = \exp(-\lambda) \times \frac{\lambda^z}{z!} \quad \forall z \in \mathbb{N} \tag{6.17}$$

Si l'on pose $\lambda = 1$, la fonction de masse devient :

$$f_Z(z) = \Pr(Z = z) = \frac{\exp(-1)}{z!} \quad \forall z \in \mathbb{N} \tag{6.18}$$

On peut calculer les probabilités associées comme suit :

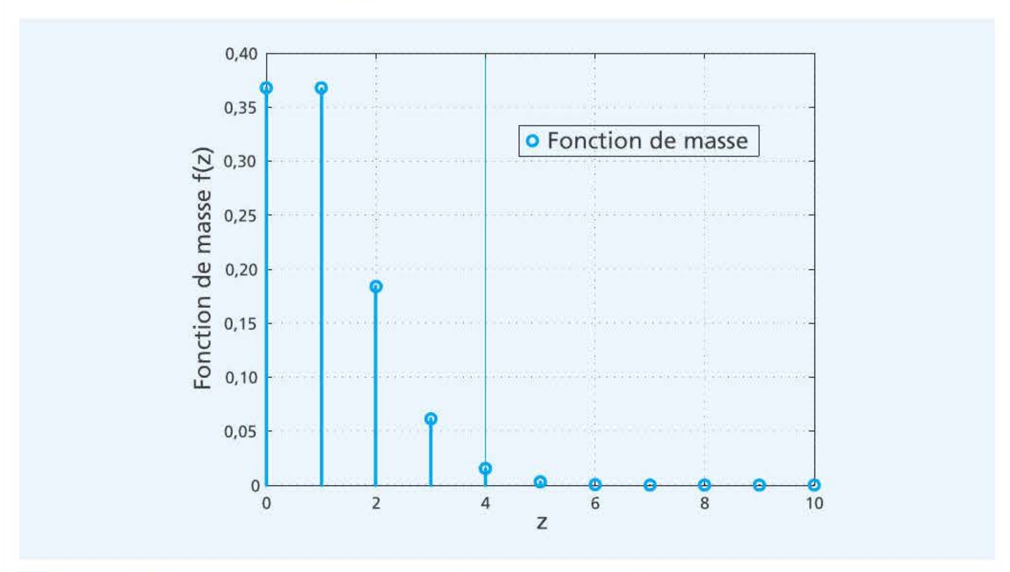
$$f_Z(0) = \Pr(Z = 0) = \frac{\exp(-1)}{0!} = \exp(-1) \quad (6.19)$$

$$f_Z(1) = \Pr(Z = 1) = \frac{\exp(-1)}{1!} = \exp(-1) \quad (6.20)$$

$$f_Z(2) = \Pr(Z = 2) = \frac{\exp(-1)}{2!} = \frac{\exp(-1)}{2} \quad (6.21)$$

$$\ldots$$

et ainsi de suite pour toutes les valeurs de z appartenant à $\mathbb{N}$. Cette fonction de masse évaluée pour les réalisations allant de 0 à 10 est représentée sur la figure 6.1. Dans ce cas, on ne peut pas représenter la loi de probabilité de Z par les valeurs des probabilités $\Pr(Z = 0)$, $\Pr(Z = 1)$, $\Pr(Z = 2)$,..., puisqu'il y en a une infinité. On représente donc cette loi de probabilité par sa *fonction de masse* (équation (6.18)).



▲ **Figure 6.1 Fonction de masse de la loi de Poisson de paramètre $\lambda = 1$**

Dans ces deux exemples, la fonction de masse dépend d'un ou de plusieurs paramètres : les paramètres n et p dans le cas de la loi binomiale (premier exemple) et le paramètre λ dans le cas de la loi de Poisson (deuxième exemple). Ces lois de probabilité sont dites paramétriques.

Définition 6.7

Une loi de probabilité **paramétrique** est associée à une fonction de masse qui dépend d'un ou de plusieurs paramètres, notés θ. On note alors cette fonction de masse sous les formes équivalentes suivantes :

$$f_X(x) \equiv f_X(x;\theta) \equiv \Pr(X = x) \equiv \Pr(X = x;\theta) \quad (6.22)$$

où le signe ≡ signifie « équivalent à ».

FOCUS

Les lois de probabilité usuelles

Certaines lois de probabilité ont des propriétés particulières et sont, pour cela, très souvent employées pour modéliser certains phénomènes de la vie quotidienne ou de la vie économique. Du fait de leur utilisation fréquente, on les qualifie de **lois (de probabilité) usuelles**. Ces lois possèdent des noms, par exemple loi *binomiale*, loi de Poisson, loi géométrique, loi binomiale négative, etc. Ce sont souvent des **lois paramétriques**. Par exemple, la fonction de masse associée à une loi de Poisson dépend d'un paramètre réel positif noté λ. La fonction de masse d'une loi binomiale dépend de deux paramètres, souvent notés n et p, tels que $n > 0$ et $p \in [0,1]$.

Ces lois usuelles sont représentées par des symboles, qui souvent sont des raccourcis de leur nom et font apparaître les paramètres de leur fonction de masse. Par exemple, la loi binomiale est notée $\mathcal{B}(n,p)$, la loi de Poisson est notée $\mathcal{P}(\lambda)$, etc. Lorsque l'on veut mentionner qu'une variable aléatoire X est distribuée selon une loi usuelle on utilise ce symbole abrégé, précédé du signe $\sim$ qui signifie « est distribué selon ». Ainsi, l'expression $X \sim \mathcal{P}(\lambda)$ signifie que la variable aléatoire X est distribuée selon une loi de Poisson de paramètre λ. Dit autrement, la loi de probabilité de X est une loi de Poisson de paramètre λ. On dit aussi que la variable aléatoire X suit une loi de Poisson de paramètre λ. Ces trois phrases sont équivalentes (▶ chapitre 7).

2.2 Fonction de répartition et quantile

La loi de probabilité d'une variable aléatoire discrète peut aussi être caractérisée par sa **fonction de répartition**[3]. Quel que soit le type de variable aléatoire (discrète ou continue), la fonction de répartition est toujours définie de la façon suivante.

Définition 6.8

La **fonction de répartition** de la variable aléatoire X, notée $F_X(x)$, correspond à la probabilité que cette variable prenne des réalisations inférieures ou égales à une certaine valeur $x \in \mathbb{R}$:

$$F_X(x) = \Pr(X \leq x) \quad \forall x \in \mathbb{R} \tag{6.23}$$

La fonction de répartition (*cumulative distribution function* ou cdf, en anglais) est toujours notée avec une lettre majuscule, par exemple $F_X(x)$, par opposition à la fonction de masse, notée avec une lettre minuscule (par exemple $f_X(x)$).

Remarque : La fonction de répartition, contrairement à la fonction de masse, est définie pour toute valeur réelle x et pas uniquement sur les valeurs des réalisations appartenant à $X(\Omega)$. Par exemple, si le support de la variable aléatoire X est $X(\Omega) = \{0,1,2\}$, on peut calculer $F_X(1{,}57) = \Pr(X \leq 1{,}57)$, $F_X(4) = \Pr(X \leq 4)$ ou même $F_X(-3{,}1) = \Pr(X \leq -3{,}1)$. C'est pourquoi cette définition de la fonction de répartition est valable tant pour les variables aléatoires discrètes, que pour les variables aléatoires continues (▶ section 3).

3 On distingue la fonction de répartition d'une variable aléatoire X de la *fonction de répartition empirique* (▶ chapitre 1) associée à une population ou un échantillon $x_1, \ldots, x_n$.

La quantité $\Pr(X \leq x)$ est appelée **probabilité cumulée**, car elle correspond au cumul (c'est-à-dire à la somme dans le cas des variables discrètes) de toutes les probabilités associées à des réalisations $x_i \in X(\Omega)$ inférieures ou égales à x.

Définition 6.9

La **fonction de répartition** d'une variable aléatoire discrète X définie sur le support $X(\Omega)$ est une fonction $F_X(x) : \mathbb{R} \rightarrow [0,1]$ telle que :

$$F_X(x) = \Pr(X \leq x) = \sum_{x_i \in X(\Omega),\ x_i \leq x} \Pr(X = x_i) \quad \forall x \in \mathbb{R} \tag{6.24}$$

Dans le cas d'une variable aléatoire discrète, la fonction de répartition est une fonction croissante des valeurs de x qui se présente en *forme de fonction en escalier.*

Exemple

On considère une variable aléatoire discrète Y distribuée selon une loi binomiale $\mathcal{B}(2{,}0{,}5)$ ayant pour support $Y(\Omega) = \{0,1,2\}$ et de fonction de masse :

$$f_Y(y_i) = \Pr(Y = y_i) = \frac{2!}{y_i! \times (2 - y_i)!} \times 0{,}5^{y_i} \times 0{,}5^{2-y_i} \quad \forall y_i \in \{0,1,2\} \tag{6.25}$$

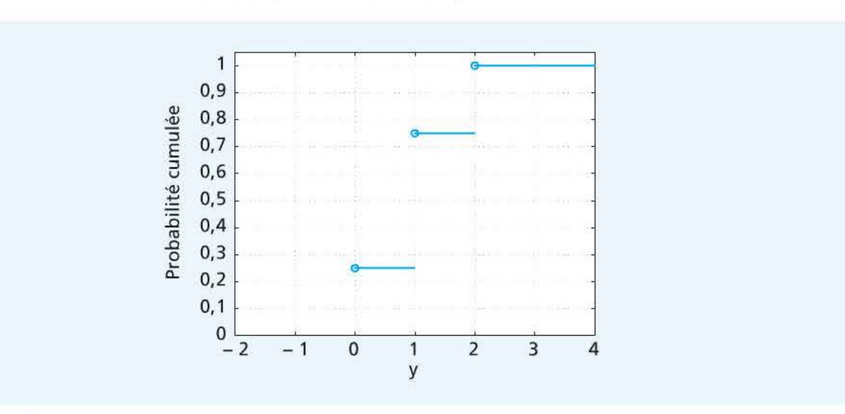
Rappelons que $\Pr(Y = 0) = 0{,}25$, $\Pr(Y = 1) = 0{,}5$ et $\Pr(Y = 2) = 0{,}25$. Déterminons sa fonction de répartition $F_Y(y)$, $\forall y \in \mathbb{R}$. Pour cela, considérons plusieurs cas :

- Si $y < 0$, alors $F_Y(y) = \Pr(Y \leq y) = 0$ puisqu'il n'existe pas de réalisation de y strictement inférieure à 0.
- Si $0 \leq y < 1$, alors $F_Y(y) = \Pr(Y \leq y) = \Pr(Y = 0) = 0{,}25$ puisque 0 est la seule valeur entre 0 (inclus) et 1 (exclu) pour laquelle une probabilité existe. Toutes les autres valeurs correspondent à des événements impossibles (probabilité nulle).
- Si $1 \leq y < 2$, alors $F_Y(y) = \Pr(Y \leq y) = \Pr(Y = 0) + \Pr(Y = 1) = 0{,}75$.
- Si $y \geq 2$, alors $F_Y(y) = \Pr(Y \leq y) = \Pr(Y = 0) + \Pr(Y = 1) + \Pr(Y = 2) = 1$.

Par conséquent, la fonction de répartition de la variable Y est définie par $\forall y \in \mathbb{R}$:

$$F_Y(y) = \begin{cases} 0 & \text{si } y < 0 \\ 0{,}25 & \text{si } 0 \leq y < 1 \\ 0{,}75 & \text{si } 1 \leq y < 2 \\ 1 & \text{si } y \geq 2 \end{cases} \tag{6.26}$$

Cette fonction de répartition est représentée sur la figure 6.2. Il convient de noter que cette fonction en escalier est *discontinue* pour les valeurs 0, 1 et 2.



▲ **Figure 6.2 Fonction de répartition de la loi binomiale $\mathcal{B}(2{,}0.5)$**

Propriété

Fonction de répartition

Pour toute variable aléatoire X (discrète ou continue), la fonction de répartition $F_X(x)$ vérifie toujours les propriétés suivantes :

1. $0 \leq F_X(x) \leq 1, \forall x \in \mathbb{R}$.
2. $\lim\limits_{x \to -\infty} F_X(x) = 0$.
3. $\lim\limits_{x \to +\infty} F_X(x) = 1$.

La première propriété signifie que la fonction de répartition en tant que probabilité cumulée, est nécessairement comprise entre $[0,1]$. La seconde propriété signifie que la probabilité cumulée que les réalisations de X soient plus petites qu'une valeur tendant vers $-\infty$, est nécessairement égale à 0, *i.e.* $F_X(-\infty) = \Pr(X \leq -\infty) = 0$. L'événement $X \leq -\infty$ est en effet un événement impossible. Parallèlement, la probabilité cumulée que les réalisations de X soient plus petites qu'une valeur tendant vers $+\infty$, est nécessairement égale à 1 : $F_X(+\infty) = \Pr(X \leq +\infty) = 1$. Dit autrement, l'événement $X \leq \infty$ est un événement certain.

Remarque : La fonction de répartition permet de calculer la probabilité que la variable aléatoire X appartienne à un **intervalle** $[a,b]$ où $(a,b) \in \mathbb{R}^2$ avec $b > a$.

$$\Pr(a \leq X \leq b) = \Pr(X \leq b) - \Pr(X \leq a) = F_X(b) - F_X(a) \qquad (6.27)$$

Nous avons vu que la fonction de répartition est une fonction qui pour toute valeur $x \in \mathbb{R}$ associe la probabilité cumulative $F_X(x) = \Pr(X \leq x)$. Il est possible « d'inverser »[4] cette fonction de répartition afin de déterminer la valeur de x qui correspond à une certaine probabilité cumulative $\alpha = \Pr(X \leq x)$, avec $\alpha \in [0,1]$. On parle alors de **fonction de répartition inverse** ou de **quantile d'ordre** α.

Définition 6.10

Le **quantile d'ordre** α de la loi de probabilité de X, noté $F_X^{-1}(\alpha)$, est la plus petite réalisation appartenant à $X(\Omega)$ associée à une probabilité cumulée supérieure ou égale à α :

$$F_X\left(F_X^{-1}(\alpha)\right) = \Pr\left(X \leq F_X^{-1}(\alpha)\right) \geq \alpha \quad \forall \alpha \in [0,1] \qquad (6.28)$$

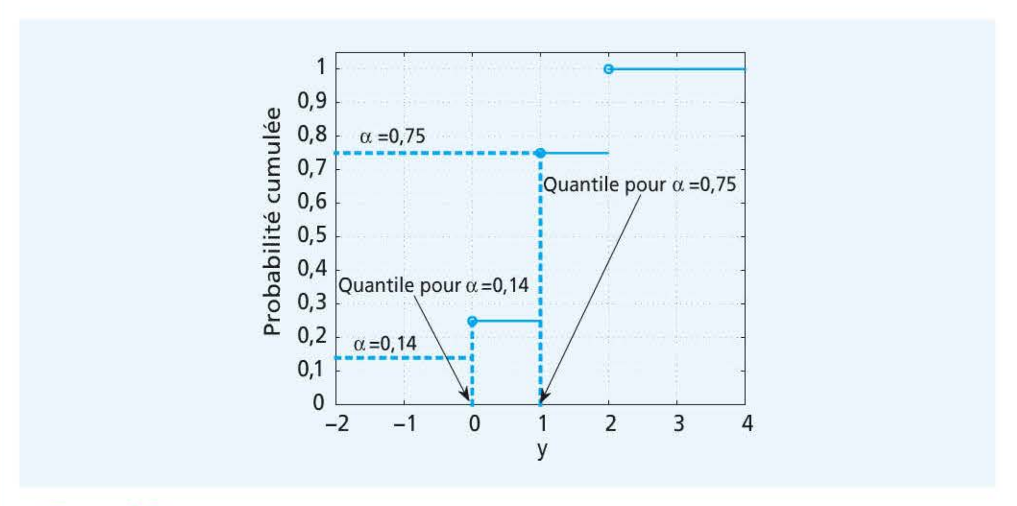
Par construction, un quantile d'une loi de distribution discrète appartient à l'univers des réalisations $X(\Omega)$: c'est une réalisation de la variable X. Le quantile d'ordre α peut être noté $F_X^{-1}(\alpha)$ ou Q_α. L'interprétation d'un quantile est la suivante. Si le quantile d'ordre $\alpha = 0{,}05$ est égal à $F_X^{-1}(0{,}05) = 2$, cela signifie qu'il y a 5 % de chances que les réalisations de la variable aléatoire discrète X soient inférieures ou égales à 2. Reprenons l'exemple précédent.

Exemple

On considère une variable aléatoire discrète Y distribuée selon une loi binomiale $\mathcal{B}(n,p)$ avec $n = 2$ et $p = 0{,}5$, ayant pour support $Y(\Omega) = \{0,1,2\}$. Sa fonction de répartition $F_Y(y)$

4 Dans le cas d'une variable aléatoire discrète, la fonction de répartition ne peut pas être inversée au sens propre.

est définie par l'équation (6.26). Déterminons les quantiles d'ordre $\alpha = 0{,}14$ et $\alpha = 0{,}75$ associés à la loi de probabilité de Y. Notons $a = F_Y^{-1}(0{,}14)$, le quantile d'ordre $\alpha = 0{,}14$. Ce quantile correspond à la plus petite réalisation de Y (c'est-à-dire soit 0, 1 ou 2) telle que la probabilité cumulée $\Pr(Y \leq a)$ soit égale ou supérieure à 0,14. On vérifie sur la figure 6.3 que cette réalisation est égale à 0, *i.e.* $F_Y^{-1}(0{,}14) = 0$. De la même façon, on vérifie que le quantile d'ordre $\alpha = 0{,}75$ est égal à 1, *i.e.* $F_Y^{-1}(0{,}75) = 1$.



▲ **Figure 6.3 Quantiles d'ordres $\alpha = 0{,}14$ et $\alpha = 0{,}75$ de la loi binomiale $\mathcal{B}(2, 0{,}5)$**

Remarque : Le quantile d'ordre $\alpha = 0{,}5$ de la loi de probabilité est appelé la **médiane**, il est peut être noté sous différentes formes équivalentes, *i.e.* $Q_2 = Q_{0.5} = F_X^{-1}(0{,}5)$. Les quantiles à 25 % et 75 % sont respectivement notés Q_1 et Q_3. La distance $Q_3 - Q_1$ est appelée **écart interquartile** ou interquartile. Cet écart est une mesure de la dispersion (▶ chapitre 1) des réalisations de la variable aléatoire.

2.3 Moments d'une variable aléatoire discrète

Une loi de probabilité discrète peut être caractérisée de façon équivalente par sa fonction de masse ou par sa fonction de répartition. Il existe une troisième façon de donner la même information : on utilise pour cela les **moments**[5]. Les moments sont des indicateurs de dispersion de la loi de probabilité. Ainsi, il est possible de définir une loi de probabilité à partir de la **population de ses moments**, c'est-à-dire à partir de l'ensemble des moments qui peuvent être associés à cette distribution.

On distingue les **moments ordinaires** (ou moments non centrés) des **moments centrés**. La définition générale des moments (ou intégrale de Riemann–Stieltjes), applicable tant dans le cas d'une variable aléatoire discrète que dans le cas d'une variable aléatoire continue, est la suivante (▶ définition 6.11).

5 Il convient de distinguer les moments (théoriques) associés à une variable aléatoire, des moments empiriques (▶ chapitre 1) associés à un échantillon ou une population.

EN PRATIQUE

Les logiciels de statistique et d'économétrie

Dans la plupart des **logiciels d'économétrie** et de **statistique**, il existe des fonctions préprogrammées qui permettent d'obtenir les valeurs de la fonction de masse, de la fonction de répartition et les quantiles de la plupart des distributions usuelles.

Par exemple dans le logiciel Matlab (éditeur MathWorks), ces fonctions s'écrivent toujours sous la même forme : une abréviation du nom de la loi (par exemple *bino* pour la loi binomiale) suivie d'un acronyme précisant le type de fonction.

L'acronyme *cdf* renvoie à la fonction de répartition (pour *cumulative distribution function*), *inv* (pour *inverse*) au fractile, *pdf* (pour *probability density function*) à la fonction de densité ou à la fonction de masse, etc.

Sur la figure 6.4 est reproduite une copie de l'aide (en anglais) de la fonction *binoinv* qui correspond à la fonction permettant de calculer le fractile d'une loi binomiale $\mathcal{B}(n,p)$. On trouve en outre sur cette figure un exemple d'appel de la fonction *binoinv* utilisée pour déterminer le fractile d'ordre $\alpha = 0{,}14$ d'une loi binomiale $\mathcal{B}(2;0{,}5)$ comme dans le cas de notre exemple. On retrouve bien sûr le résultat $F_Y^{-1}(0{,}14) = 0$.

```
>> help binoinv
 BINOINV Inverse of the binomial cumulative distribution function (cdf).
    X = BINOINV(Y,N,P) returns the inverse of the binomial cdf with
    parameters N and P. Since the binomial distribution is
    discrete, BINOINV returns the least integer X such that
    the binomial cdf evaluated at X, equals or exceeds Y.

    The size of X is the common size of the input arguments. A scalar input
    functions as a constant matrix of the same size as the other inputs.

    Note that X takes the values 0,1,2,...,N.

    See also binocdf, binofit, binopdf, binornd, binostat, icdf.

    Reference page in Help browser
       doc binoinv

>> binoinv(0.14,2,0.5)

ans =

     0
```

▲ **Figure 6.4** **Exemple d'utilisation de la fonction *binoinv* sous Matlab**

Définition 6.11

Le **moment ordinaire (non centré) d'ordre** $k \in \mathbb{N}$ de la loi de probabilité de X est défini par :

$$m_k = \mathbb{E}\left(X^k\right) = \int_{-\infty}^{+\infty} x^k dF_X(x) \tag{6.29}$$

Le **moment centré d'ordre** $k \in \mathbb{N}$ de la loi de probabilité de X est défini par :

$$\mu_k = \mathbb{E}\left((X - \mathbb{E}(X))^k\right) = \int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^k \, dF_X(x) \tag{6.30}$$

où $F_X(.)$ désigne la fonction de répartition de la loi de X.

L'expression $\mathbb{E}(X)$ se prononce « espérance de X ». La suite $(m_k)_{k\in\mathbb{N}}$ caractérise la population des moments ordinaires, c'est-à-dire l'ensemble des moments ordinaires $m_0,m_1,m_2,\ldots,m_k,\ldots$, que l'on peut définir à partir de la loi de la variable X. De même, la suite $(\mu_k)_{k\in\mathbb{N}}$ caractérise la population des moments centrés. On peut montrer que connaître la population des moments ordinaires $(m_k)_{k\in\mathbb{N}}$ ou des moments centrés $(\mu_k)_{k\in\mathbb{N}}$ est équivalent à connaître la loi de probabilité de la variable X. Par définition :

$$m_0 = \mathbb{E}(1) = 1 \qquad \mu_0 = \mathbb{E}(1) = 1 \qquad \mu_1 = \mathbb{E}(X - \mathbb{E}(X)) = 0 \tag{6.31}$$

puisque $\mathbb{E}(X - \mathbb{E}(X)) = \mathbb{E}(X) - \mathbb{E}(X) = 0$.

Dans le cas d'une *variable aléatoire discrète*, les moments peuvent s'exprimer en fonction de la fonction de masse.

Définition 6.12

Soit X une **variable aléatoire discrète** caractérisée par une fonction de masse $f_X(x) = \Pr(X = x)$, $\forall x \in X(\Omega)$ où $X(\Omega) = \{x_1,\ldots,x_n\}$ est un univers de dimension finie. Le **moment ordinaire** d'ordre $k \in \mathbb{N}$ de la loi de probabilité de X est défini par :

$$m_k = \mathbb{E}\left(X^k\right) = \sum_{i=1}^{n} x_i^k \Pr(X = x_i) \tag{6.32}$$

Le **moment centré d'ordre** $k \in \mathbb{N}$ de la loi de probabilité de X est défini par :

$$\mu_k = \mathbb{E}\left((X - \mathbb{E}(X))^k\right) = \sum_{i=1}^{n} (x_i - \mathbb{E}(X))^k \Pr(X = x_i) \tag{6.33}$$

Exemple

On considère une variable aléatoire discrète Y admettant une *distribution de Bernouilli* de paramètre $p \in [0,1]$. Le support de cette distribution $Y(\Omega) = \{0,1\}$ est fini et sa fonction de masse est définie par :

$$f_Y(y) = \Pr(Y = y) = p^y (1-p)^{1-y} \quad \forall y_i \in \{0,1\} \tag{6.34}$$

Déterminons les trois premiers moments ordinaires associés à la loi de probabilité de Y. Nous savons que $m_0 = \mathbb{E}\left(Y^0\right) = 1$. Pour les ordres suivants, il vient :

$$m_1 = \mathbb{E}(Y) = \sum_{i=0}^{1} i \times f_Y(i) \tag{6.35}$$

$$= 0 \times p^0 \times (1-p)^1 + 1 \times p^1 \times (1-p)^0 = p \tag{6.36}$$

$$m_2 = \mathbb{E}\left(Y^2\right) = \sum_{i=0}^{1} i^2 \times f_Y(i) \tag{6.37}$$

$$= 0^2 \times p^0 \times (1-p)^1 + 1^2 \times p^1 \times (1-p)^0 = p \tag{6.38}$$

Remarque : Dans le cas où l'univers des réalisations $X(\Omega)$ est un univers infini, mais dénombrable, tel que $X(\Omega) = \{x_1, \ldots, x_n, \ldots\}$, les définitions des moments ordinaires et centrés deviennent :

$$m_k = \mathbb{E}\left(X^k\right) = \lim_{n\to\infty} \sum_{i=1}^{n} x_i^k \ \Pr(X = x_i) \tag{6.39}$$

$$\mu_k = \mathbb{E}\left((X - \mathbb{E}(X))^k\right) = \lim_{n\to\infty} \sum_{i=1}^{n} (x_i - \mathbb{E}(X))^k \ \Pr(X = x_i) \tag{6.40}$$

Exemple

On considère une variable aléatoire discrète Z admettant une *distribution géométrique* de paramètre $p \in]0,1]$, notée $\mathcal{G}(p)$, telle que $Z(\Omega) = \mathbb{N}^*$. Sa fonction de masse est définie par :

$$f_Z(z_i) = \Pr(Z = z_i) = (1-p)^{z_i - 1}\, p \quad \forall z_i \in \{1, 2, \ldots, n, \ldots\} \tag{6.41}$$

Dans ce cas, l'univers de résultats $Z(\Omega)$ est infini dénombrable. Déterminons les trois premiers moments ordinaires associés à la loi de probabilité de Z. Nous savons que $m_0 = \mathbb{E}\left(Z^0\right) = 1$. Pour les ordres suivants, il vient :

$$m_1 = \mathbb{E}(Z) = \lim_{n\to\infty} \sum_{i=1}^{n} i \times (1-p)^{i-1} \times p = \frac{1}{p} \tag{6.42}$$

$$m_2 = \mathbb{E}\left(Z^2\right) = \lim_{n\to\infty} \sum_{i=1}^{n} i^2 \times (1-p)^{i-1} \times p = \frac{2-p}{p^2} \tag{6.43}$$

Pour retrouver ces résultats, il convient d'appliquer la notion de série entière à $(1-p)$ et d'utiliser l'identité géométrique.

Remarque : Pour certaines lois de probabilité, certains moments ordinaires ou centrés *n'existent pas*. On dit que le moment ordinaire d'ordre k *n'existe pas* lorsque :

$$\mathbb{E}\left(\left|X^k\right|\right) = \int_{-\infty}^{+\infty} \left|x^k\right| dF_X(x) = +\infty \tag{6.44}$$

L'exemple précédent nous a montré que, dans le cas d'une loi à support infini, le calcul des moments fait généralement appel à la résolution de suites. Une façon plus simple de retrouver les moments consiste à utiliser la **fonction génératrice des moments**.

Définition 6.13

La **fonction génératrice des moments** d'une variable aléatoire X, telle que $\mathbb{E}(X)$ existe, est définie par :

$$M_X(t) = \mathbb{E}(\exp(tX)) = \int_{-\infty}^{+\infty} \exp(tx)\ dF_X(x) \qquad \forall t \in \mathbb{R} \tag{6.45}$$

De cette définition générale, valable pour tout type de variable (discrète ou continue), nous pouvons déduire une définition spécifique aux variables discrètes.

Définition 6.14

Soit X une *variable aléatoire discrète* définie sur un univers fini $X(\Omega) = \{x_1, \ldots, x_n\}$, sa **fonction génératrice des moments** est égale à :

$$M_X(t) = \mathbb{E}(\exp(tX)) = \sum_{i=1}^{n} \exp(tx_i)\ \Pr(X = x_i) \tag{6.46}$$

Quel est le lien entre la fonction génératrice des moments et les moments ordinaires $(m_k)_{k\in\mathbb{N}}$? On peut montrer que la fonction génératrice des moments peut toujours se réécrire sous la forme d'un développement en séries entières tel que :

$$M_X(t) = 1 + tm_1 + \frac{t^2 m_2}{2!} + \frac{t^3 m_3}{3!} + \ldots \tag{6.47}$$

En dérivant cette fonction par rapport à t, on obtient :

$$M_X'(t) = \frac{\partial M_X(t)}{\partial t} = m_1 + tm_2 + \frac{t^2 m_3}{2} + \ldots \tag{6.48}$$

Si l'on souhaite obtenir le moment ordinaire d'ordre 1, il suffit d'évaluer l'expression de cette dérivée en $t = 0$, on obtient immédiatement $M_X'(0) = m_1$. Si l'on souhaite obtenir le moment d'ordre 2, il convient alors de dériver deux fois la fonction génératrice et d'évaluer cette dérivée seconde en 0, *i.e.* $M_X''(0) = m_2$, et ainsi de suite.

Propriété

Fonction génératrice des moments

Si le moment ordinaire d'ordre $k \in \mathbb{N}$ de la variable X existe, il correspond à la dérivée $k^{ème}$ de la fonction génératrice des moments évaluée au point $t = 0$.

$$m_k = \left.\frac{\partial^k M_X(t)}{\partial t^k}\right|_{t=0} \tag{6.49}$$

Le signe ∂ correspond à la notion de dérivée partielle. L'expression $\partial^k f(x)/\partial x^k$, avec un exposant k sur le signe ∂ au numérateur et un exposant k sur la variable de dérivation x au dénominateur, correspond donc à la dérivée $k^{ème}$ de la fonction $f(x)$. La barre verticale signifie « évaluée en ». Ainsi, $\partial^k f(x)/\partial x^k|_{x_0}$ correspond à la dérivée $k^{ème}$ de la fonction $f(x)$ évaluée au point x_0.

Exemple

On considère une variable aléatoire discrète Z admettant une *distribution géométrique* de paramètre $p \in]0,1]$, notée $\mathcal{G}(p)$, telle que $Z(\Omega) = \mathbb{N}^*$. On admet que sa fonction génératrice des moments est définie par :

$$M_Z(t) = \frac{pe^t}{1 - qe^t} \tag{6.50}$$

avec $q = 1 - p$. Déterminons les moments ordinaires d'ordres 1 et 2 associés à la loi de probabilité de Z. Commençons par dériver la fonction génératrice des moments :

$$\frac{\partial M_Z(t)}{\partial t} = \frac{pe^t}{(1 - qe^t)^2} \qquad \frac{\partial^2 M_Z(t)}{\partial t^2} = \frac{pe^t\left(1 + 2q - 2\left(q + q^2\right)e^t + q^2 e^{2t}\right)}{(1 - qe^t)^3} \tag{6.51}$$

Dès lors, il vient :

$$m_1 = \mathbb{E}(Z) = \left.\frac{\partial M_X(t)}{\partial t}\right|_{t=0} = \frac{pe^0}{(1 - qe^0)^2} = \frac{p}{(1 - (1 - p))^2} = \frac{1}{p} \tag{6.52}$$

$$m_2 = \mathbb{E}\left(Z^2\right) = \left.\frac{\partial^2 M_X(t)}{\partial t^2}\right|_{t=0} = \frac{pe^0\left(1 + 2q - 2\left(q + q^2\right)e^0 + q^2 e^0\right)}{(1 - qe^0)^3} \tag{6.53}$$

$$= \frac{p\left(1 + 2q - 2\left(q + q^2\right) + q^2\right)}{(1 - q)^3} = \frac{2 - p}{p^2} \tag{6.54}$$

On vérifie que l'on retrouve les expressions de m_1 et m_2 obtenues précédemment.

2.4 Moments remarquables

Certains des moments sont si importants que l'on leur a attribué un nom spécifique : *espérance*, *variance*, etc. Ces sont les **moments remarquables**.

2.4.1 Espérance

L'**espérance**, notée $\mathbb{E}(X)$ ou $E(X)$, correspond au moment ordinaire d'ordre un, *i.e.* m_1. Ce moment donne une idée de la « moyenne »[6] des réalisations de la variable aléatoire X que l'on peut obtenir si l'on effectue plusieurs tirages de cette variable. Plus précisément, l'espérance est définie comme la somme pondérée des réalisations dans laquelle les pondérations sont déterminées par les probabilités associées.

Définition 6.15

L'**espérance** d'une *variable aléatoire discrète* X définie sur un support fini $X(\Omega) = \{x_1, \ldots, x_n\}$ est égale à :

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i \ \Pr(X = x_i) \tag{6.55}$$

Dans le cas d'un support $X(\Omega)$ infini dénombrable, cette définition devient :

$$\mathbb{E}(X) = \lim_{n \to \infty} \sum_{i=1}^{n} x_i \ \Pr(X = x_i) \tag{6.56}$$

Exemple

On considère une variable aléatoire Y définie sur $Y(\Omega) = \{0,2,4,6\}$ telle que sa loi de probabilité est caractérisée par les probabilités du tableau 6.3.

▼ Tableau 6.3 Probabilités associées à la variable *Y*

Réalisation de *Y*	Probabilité
$Y = 0$	$\Pr(Y = 0) = 0{,}1$
$Y = 2$	$\Pr(Y = 2) = 0{,}3$
$Y = 4$	$\Pr(Y = 4) = 0{,}4$
$Y = 6$	$\Pr(Y = 6) = 0{,}2$

Son espérance est égale à :

$$\mathbb{E}(Y) = \sum_{i=1}^{4} y_i \ \Pr(Y = y_i) = 0 \times 0{,}1 + 2 \times 0{,}3 + 4 \times 0{,}4 + 6 \times 0{,}2 = 3{,}4 \tag{6.57}$$

Cela signifie qu'en « moyenne » les réalisations obtenues pour plusieurs tirages dans la loi de probabilité de cette variable seront égales à 3,4.

6 Il ne faut surtout pas confondre les concepts de moyenne empirique (▶ chapitre 1) et d'espérance. Comme nous le verrons dans le chapitre 9 consacré à l'estimation, la moyenne empirique est une variable aléatoire (un estimateur) alors que l'espérance est une constante.

Exemple

On considère une variable aléatoire discrète Z distribuée selon une *loi uniforme discrète* sur $Z(\Omega) = \{1,\ldots,n\}$. Sa fonction de masse reflète la propriété d'*équiprobabilité* de cette distribution :

$$f_Z(z) = \Pr(Z = z) = \frac{1}{n} \quad \forall z \in Z(\Omega) \tag{6.58}$$

L'espérance de la variable Z est égale à :

$$\mathbb{E}(Z) = \sum_{i=1}^{n} i \Pr(Z = i) = \frac{1}{n}\sum_{i=1}^{n} i = \frac{1}{n} \times \frac{n(n+1)}{2} = \frac{n+1}{2} \tag{6.59}$$

Souvent dans la pratique, on est amené à exprimer une variable aléatoire *en fonction* d'une autre. Par exemple, on s'intéresse à la variable Y définie par $Y = X^2$ ou $Y = 2 + 3X$. Comment déterminer l'espérance d'une transformée ou d'une fonction de la variable X sans nécessairement connaître sa loi de probabilité ? On utilise pour ce faire la propriété suivante, valable tant pour les variables aléatoires discrètes que pour les variables aléatoires continues.

Propriété

Espérance d'une fonction de variable aléatoire

Soit X une variable aléatoire discrète définie sur un support $X(\Omega) = \{x_1,\ldots,x_n\}$ fini et soit $g(.)$ une fonction telle que $\sum_{i=1}^{n} |g(x_i)| \Pr(X = x_i) < \infty$. L'espérance de la variable aléatoire $g(X)$ est alors définie par :

$$\mathbb{E}(g(X)) = \sum_{i=1}^{n} g(x_i) \Pr(X = x_i) \tag{6.60}$$

Dans le cas d'une transformation linéaire $g(X) = a + bX$, cette propriété illustre le fait que l'espérance est un **opérateur linéaire** puisque l'on a $\mathbb{E}(g(X)) = g(\mathbb{E}(X))$.

Propriété

Linéarité de l'espérance

Soit X une variable aléatoire et soient deux constantes $(a,b) \in \mathbb{R}^2$, alors $\mathbb{E}(a + bX) = a + b\mathbb{E}(X)$. On dit que l'espérance est un opérateur linéaire.

Mais attention, il est important de noter que le résultat selon lequel $\mathbb{E}(g(X)) = g(\mathbb{E}(X))$ n'est valable que dans le cas où la fonction $g(.)$ est linéaire.

Propriété

Espérance de fonction non-linéaire

Soit X une variable aléatoire et soit $g(.)$ une fonction non-linéaire, alors :

$$\mathbb{E}(g(X)) \neq g(\mathbb{E}(X))$$

Exemple

On considère une variable aléatoire Y telle que $\mathbb{E}(Y) = 3$. Puisque l'espérance est un opérateur linéaire $\mathbb{E}(2 - 4Y) = 2 - 4\mathbb{E}(Y) = -10$. En revanche $\mathbb{E}\left(Y^2\right) \neq \mathbb{E}(Y)^2$ et $\mathbb{E}(1/Y) \neq 1/\mathbb{E}(Y)$ car les fonctions $g(x) = x^2$ et $g(x) = 1/x$ ne sont pas des fonctions linéaires.

2.4.2 Variance

La **variance**[7] est un indicateur de la dispersion de la loi de probabilité autour de l'espérance. La définition générale de la variance (cas des variables aléatoires continues ou discrètes) est la suivante.

Définition 6.16

Soit X une variable aléatoire telle que $\mathbb{E}\left(X^2\right)$ existe. La **variance** de X est définie par :

$$\mathbb{V}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) \tag{6.61}$$

La variance correspond au moment centré d'ordre 2, *i.e.* μ_2. Elle peut être interprétée comme la « *moyenne* » des réalisations en écart à l'espérance $\mathbb{E}(X)$: plus la variance est élevée, plus les réalisations de la variable X auront de grandes chances d'être éloignées de la valeur de l'espérance.

En développant l'expression de la variance, il vient :

$$\mathbb{V}(X) = \mathbb{E}\left(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2\right) \tag{6.62}$$

Puisque les quantités $\mathbb{E}(X)$ et $\mathbb{E}(X)^2$ sont des constantes et que l'espérance est un opérateur linéaire, cette expression peut se réécrire comme :

$$\mathbb{V}(X) = \mathbb{E}\left(X^2\right) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2 = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2 \tag{6.63}$$

Propriété

Formule de König-Huygens

Si $\mathbb{E}(X^2)$ existe, la variance $\mathbb{V}(X)$ peut toujours se réécrire sous la forme suivante, dite formule de König-Huygens :

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \tag{6.64}$$

Cette expression se lit comme « *l'espérance de la variable X^2 moins le carré de l'espérance de la variable X* ».

Dans le cas d'une *variable aléatoire discrète*, la variance peut être définie à partir de sa fonction de masse $f_X(x) = \Pr(X = x)$.

Définition 6.17

Soit X une variable aléatoire discrète définie sur $X(\Omega) = \{x_1, \ldots, x_n\}$. Si $\mathbb{E}(X^2)$ existe, sa variance est :

$$\mathbb{V}(X) = \sum_{i=1}^{n} (x_i - \mathbb{E}(X))^2 \Pr(X = x_i) \tag{6.65}$$

En développant l'expression de la variance de l'équation (6.65), on montre que celle-ci peut se réécrire de façon équivalente sous les formes suivantes :

$$\mathbb{V}(X) = \sum_{i=1}^{n} p_i x_i^2 - \mathbb{E}(X)^2 = \sum_{i=1}^{n} p_i x_i^2 - \left(\sum_{i=1}^{n} p_i x_i\right)^2 \tag{6.66}$$

7 On ne doit pas confondre la variance d'une variable aléatoire, notée $\mathbb{V}(X)$, et la variance empirique d'un échantillon, notée $V(X)$ (▶ chapitre 1).

avec $p_i = \Pr(X = x_i)$. Dans le cas d'un univers $X(\Omega)$ infini dénombrable, ces définitions deviennent :

$$\mathbb{V}(X) = \lim_{n\to\infty} \sum_{i=1}^{n} p_i x_i^2 - \mathbb{E}(X)^2 = \lim_{n\to\infty} \sum_{i=1}^{n} p_i x_i^2 - \left(\lim_{n\to\infty} \sum_{i=1}^{n} p_i x_i \right)^2 \quad (6.67)$$

Exemple

On considère une variable aléatoire Y définie sur $Y(\Omega) = \{0,2,4,6\}$ telle que sa loi de probabilité est caractérisée par les probabilités du tableau 6.3. Nous avons vu que son espérance était égale à $\mathbb{E}(Y) = 3{,}4$. Sa variance est égale à :

$$\mathbb{V}(Y) = \sum_{i=1}^{4} (y_i - \mathbb{E}(X))^2 \ \Pr(Y = y_i) \quad (6.68)$$

On obtient alors :

$$\mathbb{V}(Y) = (0 - 3{,}4)^2 \times 0{,}1 + (2 - 3{,}4)^2 \times 0{,}3 + (4 - 3{,}4)^2 \times 0{,}4 + (6 - 3{,}4)^2 \times 0{,}2 = 3{,}24 \quad (6.69)$$

On peut vérifier que les 3 formules des équations (6.65) et (6.66) donnent la même valeur de la variance.

Exemple

On considère une variable aléatoire discrète Z distribuée selon une loi uniforme discrète sur $Z(\Omega) = \{1,\ldots,n\}$ telle que $f_Z(z) = \Pr(Z = z) = 1/n$. Nous savons que l'espérance de la variable Z est égale à $\mathbb{E}(Z) = (n+1)/2$. On admet que son moment ordinaire d'ordre deux est égal à :

$$\mathbb{E}\left(Z^2\right) = \sum_{i=1}^{n} i^2 \ \Pr(Z = i) = \frac{1}{n} \sum_{i=1}^{n} i^2 = \frac{2n^2 + 3n + 1}{6}$$

D'après la formule de König-Huygens, la variance est égale à :

$$\mathbb{V}(Z) = \mathbb{E}\left(Z^2\right) - \mathbb{E}(Z)^2 = \frac{2n^2 + 3n + 1}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2 - 1}{12} \quad (6.70)$$

De la même façon que pour l'espérance, nous pouvons déterminer la variance d'une fonction de la variable aléatoire X, sans nécessairement connaître la loi de probabilité de cette variable transformée. Mais contrairement à l'espérance, qui est un opérateur linéaire, la variance est un **opérateur quadratique**.

Propriété

Non-linéarité de la variance

Soit X une variable aléatoire et soient deux constantes $(a,b) \in \mathbb{R}^2$, alors $\mathbb{V}(a + bX) = b^2 \mathbb{V}(X)$.

Cette propriété signifie que le fait de déplacer simplement une loi de probabilité (en ajoutant a) ne modifie pas sa variance. Par contre, changer l'échelle (multiplier par b) modifie la variance. Dans ce cas, la déformation de la variance est quadratique (multipliée par b^2).

Exemple

On considère une variable aléatoire Y telle que $\mathbb{V}(Y) = 2$. Puisque la variance est un opérateur quadratique, $\mathbb{V}(2 - 4Y) = 16\mathbb{V}(Y) = 32$.

Une mesure de dispersion équivalente à la variance est donnée par l'écart-type.

Définition 6.18

L'**écart-type**, noté σ_X, correspond à la racine carrée de la variance, *i.e.* $\sigma_X^2 = \mathbb{V}(X)$.

3 Variables aléatoires continues

On considère à présent le cas où le support (ou univers des réalisations) $X(\Omega)$ de la loi de probabilité de la variable aléatoire X est non dénombrable (▸ chapitre 5). On dit alors que la variable aléatoire est **continue**. C'est le cas notamment de toutes les variables aléatoires réelles (v.a.r. en abrégé) pour lesquelles $X(\Omega) = \mathbb{R}$ ou $X(\Omega)$ correspond à une partie de l'ensemble des réels $\mathbb{R}$, par exemple $X(\Omega) =]-\infty,a]$ ou $X(\Omega) = [a,b]$ avec $(a,b) \in \mathbb{R}^2$. La définition formelle[8] d'une **variable aléatoire réelle** (continue) est la suivante.

Définition 6.19

Soit $(\Omega,\mathcal{F},\Pr)$ un univers probabilisé non dénombrable. On appelle **variable aléatoire réelle** (continue) X toute application mesurable $X : \Omega \rightarrow X(\Omega) \subseteq \mathbb{R}$ telle que pour tout intervalle $I \subseteq X(\Omega)$:

$$\Pr(X \in I) = \Pr(\{\omega \in \Omega \,;\, X(\omega) \in I\}) \qquad (6.71)$$

Le symbole $\subseteq$, utilisé pour les ensembles, signifie « inclus ou équivalent à ». Cette définition signifie que la probabilité que la variable X appartienne à un certain intervalle de réalisations $I \subseteq X(\Omega)$ (par exemple l'intervalle $[-2,3]$) correspond à la somme des probabilités associées à tous les événements ω de l'univers Ω qui correspondent à des réalisations $X(\omega)$ qui appartiennent elles-mêmes à l'intervalle I. Ainsi, l'application mesurable X permet de déterminer les probabilités associées à des **intervalles de réalisations**.

3.1 Fonction de densité

Dans le cas d'une variable aléatoire réelle (continue), on ne peut déterminer que la probabilité associée à des intervalles de réalisations. En effet, comme nous l'avons vu dans le chapitre 5, pour une variable aléatoire continue, la probabilité d'être en un point est nulle. On dit qu'il n'y a pas de masse ponctuelle dans la densité. Par conséquent, le concept de fonction de masse n'*existe pas* pour les variables aléatoires continues.

Propriété

Variable aléatoire continue

La probabilité associée à une réalisation particulière d'une variable aléatoire continue est nulle :

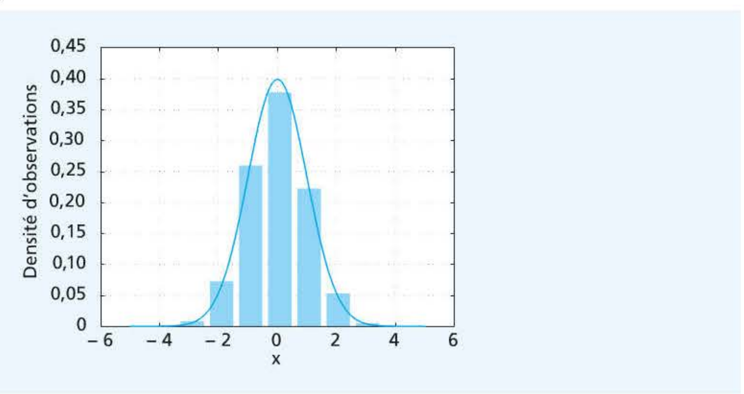
$$\Pr(X = x) = 0 \quad \forall x \in X(\Omega) \qquad (6.72)$$

8 On trouve aussi parfois la définition équivalente $\forall I \in X(\Omega)$, $X^{-1}(I) = \{\omega \in \Omega \,;\, X(\omega) \in I\} \in \mathcal{F}$, qui correspond à la définition d'une application mesurable.

Dès lors, comment représenter la loi de probabilité (ou distribution) d'une variable aléatoire continue ? On utilise pour cela le concept de **fonction de densité** (*probability density function* ou pdf en anglais).

Afin de comprendre l'intuition d'une fonction de densité, imaginons une variable aléatoire X à valeurs sur $\mathbb{R}$ admettant une loi de probabilité continue. Supposons que l'on effectue un très grand nombre de tirages dans cette loi de probabilité, par exemple 1 million. À partir de ce million de réalisations on construit un histogramme (▶ chapitre 1) comme celui représenté sur la figure 6.5. Rappelons que pour chaque classe (intervalle) de valeurs sur $\mathbb{R}$, l'histogramme indique la fréquence des réalisations appartenant à cette classe, c'est-à-dire le nombre de réalisations appartenant à cette classe divisé par le nombre total de réalisations.

Imaginons maintenant une fonction hypothétique qui, pour toutes les valeurs admissibles des réalisations (axe des abscisses), renverrait la valeur du sommet de classe de l'histogramme auquel appartient cette réalisation (axe des ordonnées), c'est-à-dire la fréquence d'observation de cette classe. Pour un nombre de classes fini cette fonction serait discontinue, sous forme de plateaux ou d'escaliers. Comme l'illustre la figure 6.5, la fonction de densité peut être interprétée comme cette fonction qui relierait les sommets de l'histogramme, dans le cas impossible[9] où le nombre de classes de l'histogramme tendrait vers l'infini ou de façon équivalente, lorsque l'amplitude des classes tendrait vers 0.



▲ **Figure 6.5** **Interprétation de la fonction de densité**

Définition 6.20

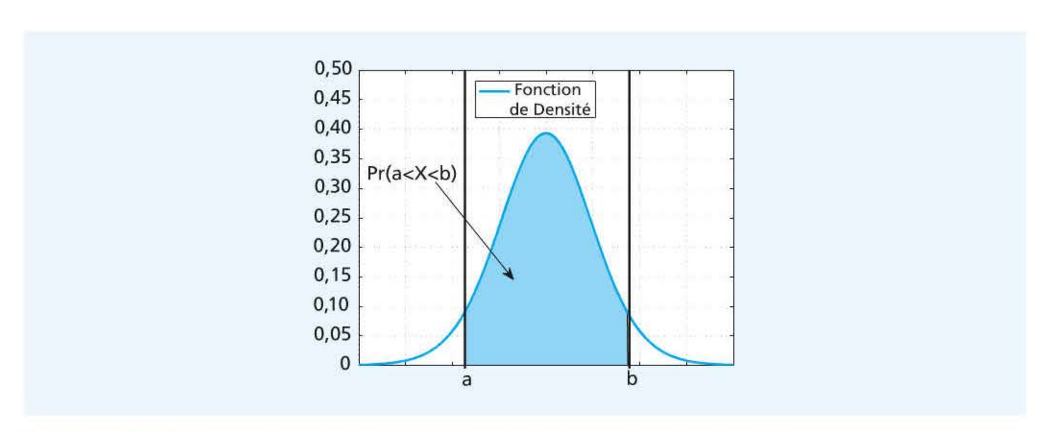
Soit X une variable aléatoire réelle définie sur le support $X(\Omega) \subseteq \mathbb{R}$. La loi de probabilité de X admet une **fonction de densité**, notée $f_X(x)$, si cette fonction est définie sur $X(\Omega)$, positive ou nulle, intégrable[10] et telle que $\forall (a,b) \in X(\Omega)^2$:

$$\Pr(a \leq X \leq b) = \int_a^b f_X(x)\, dx \qquad (6.73)$$

9 Ce cas est impossible car si le nombre de classes de l'histogramme tend vers l'infini, la fréquence de chaque classe tend nécessairement vers 0 pour un nombre de tirages donné.

10 Pour rappel, on dit qu'une fonction est *intégrable* si cette fonction peut être intégrée et que son intégrale est égale à une quantité finie.

La convention de notation pour la fonction de densité est la même que pour la fonction de masse. La fonction de densité est notée avec une lettre minuscule avec en indice le nom de la variable aléatoire (notée en majuscule). La figure 6.6 illustre le concept de fonction de densité. Une fonction de densité $f_X(x)$ est une fonction positive ou nulle, définie sur $X(\Omega) \subseteq \mathbb{R}$ (axe des abscisses) telle que pour tout couple de valeurs $(a,b) \in X(\Omega)^2$, la probabilité que les réalisations de X appartiennent à l'intervalle $[a,b]$ correspond à l'*aire sous la densité* entre ces deux bornes. Rappelons que cette aire représente l'intégrale $\int_a^b f_X(x)\,dx$.



▲ **Figure 6.6 Illustration de la définition de la fonction de densité**

Remarque : Puisque la probabilité d'être en un point est nulle, la définition de la densité peut se réécrire de façon équivalente sous les formes suivantes :

$$\Pr(a < X < b) = \Pr(a < X \leq b) = \Pr(a \leq X < b) = \int_a^b f_X(x)\,dx \qquad (6.74)$$

Dans le tableau 6.4 sont reportés quelques exemples de fonctions de densité associées à des *lois continues usuelles* (▶ focus : les lois usuelles).

▼ **Tableau 6.4 Exemples de fonctions de densité**

Nom de la loi	Fonction de densité	Paramètres	Support
Uniforme (continue)	$f_X(x) = \frac{1}{d-c}$	$(d,c) \in \mathbb{R}^2$	$X(\Omega) = [c,d]$
Exponentielle	$f_X(x) = \lambda \exp(-\lambda x)$	$\lambda > 0$	$X(\Omega) = \mathbb{R}^+$
Normale	$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$	$\mu \in \mathbb{R}, \sigma > 0$	$X(\Omega) = \mathbb{R}$
Normale standard	$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$	aucun	$X(\Omega) = \mathbb{R}$

Remarque : On constate que la plupart de ces fonctions de densité dépendent d'un ou de plusieurs paramètres. Comme pour les variables discrètes, ces densités correspondent à des **lois (continues) paramétriques**, pour lesquelles la fonction de densité peut être notée $f_X(x;\theta)$ où θ désigne l'ensemble des paramètres.

Exemple

On suppose que la note Y d'un étudiant à son examen de statistique a été attribuée au hasard. Formellement, ceci revient à supposer que la variable Y définie sur $Y(\Omega) = [0,20]$ admet une loi uniforme (continue) telle que :

$$f_Y(y) = \frac{1}{20} \quad \forall y \in [0,20] \tag{6.75}$$

Déterminons la probabilité d'obtenir une note comprise entre 8 et 13, ainsi que la probabilité d'obtenir une note supérieure à 15. Par définition, il vient :

$$\Pr(8 \leq Y \leq 13) = \int_8^{13} f_X(x)\,dx = \int_8^{13} \frac{1}{20}dx = \left[\frac{x}{20}\right]_8^{13} = \frac{13}{20} - \frac{8}{20} = \frac{1}{4} \tag{6.76}$$

Si la note a été attribuée au hasard, on vérifie qu'il y a 1 chance sur 4 d'obtenir une notre comprise entre 8 et 13. De la même façon :

$$\Pr(Y \geq 15) = \int_{15}^{20} f_X(x)\,dx = \int_{15}^{20} \frac{1}{20}dx = \left[\frac{x}{20}\right]_{15}^{20} = \frac{20}{20} - \frac{15}{20} = \frac{1}{4} \tag{6.77}$$

Remarquons que dans ce cas la borne supérieure de l'intégrale est égale à 20, puisque la fonction de densité n'est définie que sur le support $Y(\Omega) = [0,20]$ de la loi de Y.

Afin de simplifier les notations, on supposera que la fonction de densité « est définie en dehors du support de la loi de X » et qu'elle prend alors une valeur nulle :

$$f_X(x) = 0 \quad \forall x \notin X(\Omega) \tag{6.78}$$

Par exemple, la densité de la loi uniforme définie sur $X(\Omega) = [c,d]$ devient :

$$f_X(x) = \begin{cases} 1/(d-c) & \forall x \in [c,d] \\ 0 & \text{sinon} \end{cases} \tag{6.79}$$

L'avantage de cette écriture c'est que dans le cas où le support de la loi de X est une partie de l'ensemble $\mathbb{R}$, *i.e.* $X(\Omega) \subset \mathbb{R}$, nous pouvons toujours écrire les probabilités d'être « *supérieur* » ou « *inférieur* » à une certaine valeur avec des $\pm\infty$ sur les bornes de l'intégrale. Par exemple, si $X(\Omega) = [0,20]$, il vient :

$$\Pr(X \geq 15) = \int_{15}^{+\infty} f_X(x)\,dx = \int_{15}^{20} f_X(x)\,dx + \int_{20}^{+\infty} 0\,dx = \int_{15}^{20} f_X(x)\,dx \tag{6.80}$$

Par conséquent, en posant $f_X(x) = 0$, $\forall x \notin X(\Omega)$, nous pouvons adopter des notations identiques (avec des $+\infty$ ou des $-\infty$ sur les bornes des intégrales) pour définir les probabilités dans le cas où $X(\Omega) = \mathbb{R}$ et dans le cas où $X(\Omega) \subset \mathbb{R}$.

Propriété

Fonction de densité

Soit X une variable aléatoire réelle continue avec $X(\Omega) \subseteq \mathbb{R}$. Sa fonction de densité $f_X(x)$ vérifie les propriétés suivantes :

1. $f_X(x) \geq 0$, $\forall x \in X(\Omega)$ et $f_X(x) = 0$, $\forall x \notin X(\Omega)$.
2. $\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$.
3. $\Pr(X \geq a) = \Pr(X > a) = \int_a^{+\infty} f_X(x)\,dx$.
4. $\Pr(X \leq b) = \Pr(X < b) = \int_{-\infty}^{b} f_X(x)\,dx$.

La première propriété signifie que la densité est toujours positive ou nulle sur le support de la loi de X, et nulle en dehors de ce support. La seconde propriété signifie que lorsque l'on intègre une fonction de densité sur son support $X(\Omega)$, cette intégrale est nécessairement égale à 1. En effet, par définition une réalisation appartient toujours à l'univers des réalisations. Par conséquent l'événement $x \in X(\Omega)$ est certain et sa probabilité est égale à 1 :

$$\Pr(x \in X(\Omega)) = \int_{x \in X(\Omega)} f_X(x)\, dx + \int_{x \notin X(\Omega)} 0\, dx = \int_{-\infty}^{+\infty} f_X(x)\, dx = 1 \quad (6.81)$$

Remarque : Une densité n'est pas une probabilité . Une fonction de densité peut être supérieure à 1 pour certaines valeurs de $x \in X(\Omega)$.

Exemple

On suppose que la variable aléatoire réelle $Z \in \mathbb{R}$ est distribuée selon une *loi normale* d'espérance $\mu = 0$ et de variance égale à $\sigma^2 = 0{,}01$ telle que :

$$f_Z(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2\sigma^2}\right) \quad \forall z \in \mathbb{R} \quad (6.82)$$

Cette fonction de densité évaluée au point $z = 0$ est supérieure à 1 puisque :

$$f_Z(0) = \frac{\exp(0)}{0{,}1 \times \sqrt{2\pi}} = \frac{10}{\sqrt{2\pi}} \simeq 3{,}9894 \quad (6.83)$$

3.2 Fonction de répartition et quantile

La loi de probabilité d'une variable aléatoire continue peut aussi être caractérisée par sa **fonction de répartition**.

Définition 6.21

La **fonction de répartition**, notée $F_X(x)$, de la variable aléatoire réelle X définie sur $X(\Omega) \subseteq \mathbb{R}$ correspond à la probabilité que cette variable soit inférieure où égale à une certaine valeur $x \in \mathbb{R}$:

$$F_X(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f_X(u)\, du \quad \forall x \in \mathbb{R} \quad (6.84)$$

On remarque que la fonction de répartition d'une variable réelle est définie pour toute valeur de $\mathbb{R}$ et pas uniquement pour des valeurs appartenant au support $X(\Omega)$, y compris lorsque ce support n'est qu'une partie de $\mathbb{R}$ (► section 2.2).

Exemple

On considère une variable aléatoire réelle Y distribuée selon une *loi continue uniforme* sur $Y(\Omega) = [c,d]$ avec $d > c$, de densité :

$$f_Y(y) = \begin{cases} 1/(d-c) & \forall y \in [c,d] \\ 0 & \text{sinon} \end{cases} \quad (6.85)$$

Déterminons sa fonction de répartition. Par définition, $\forall y \in [c,d]$:

$$F_Y(y) = \int_{-\infty}^{y} f_Y(u)\, du = \int_{-\infty}^{c} 0\, du + \int_{c}^{y} f_Y(u)\, du \quad (6.86)$$

$$= \int_{c}^{y} f_Y(u)\, du = \left[\frac{u}{d-c}\right]_c^y = \frac{y-c}{d-c} \quad (6.87)$$

Pour une valeur $y > d$, on a :

$$F_Y(y) = \int_{-\infty}^{y} f_Y(u)\,du = \int_{-\infty}^{c} 0\,du + \int_{c}^{d} f_Y(u)\,du + \int_{d}^{y} 0\,du = \int_{c}^{d} f_Y(u)\,du \qquad (6.88)$$

Or, par définition de la fonction de densité, $\int_{c}^{d} f_Y(u)\,du = 1$. Ainsi $\forall y > d,\ F_Y(y) = 1$. De la même façon, pour toute valeur $y < c$, on a :

$$F_Y(y) = \int_{-\infty}^{y} f_Y(u)\,du = \int_{-\infty}^{y} 0\,du = 0 \qquad (6.89)$$

Par conséquent, la fonction de répartition de la variable Y est définie pour toute valeur $y \in \mathbb{R}$ par :

$$F_Y(y) = \begin{cases} 0 & \text{si } y < c \\ \dfrac{y-c}{d-c} & \forall y \in [c,d] \\ 1 & \text{si } y > d \end{cases} \qquad (6.90)$$

Par exemple si la variable est définie sur $Y(\Omega) = [2,4]$, on obtient $F_Y(y) = (y-2)/2$ si $y \in [2,4]$. On peut alors calculer la probabilité que les réalisations de la variable Y soient inférieures à 3 comme $\Pr(Y \leq 3) = F_Y(3) = 1/2$, mais aussi la probabilité que les réalisations de la variable Y soient inférieures à 30 (valeur n'appartenant pas au support de la loi de Y), puisque $\Pr(Y \leq 30) = F_Y(30) = 1$.

Remarque : Dans la majorité des cas, on se contente d'exprimer la fonction de répartition sur *le support de la loi* $X(\Omega) \subseteq \mathbb{R}$.

Exemple

On considère une variable aléatoire réelle Z à valeurs sur $Z(\Omega) = \mathbb{R}^+$ et distribuée selon une *loi exponentielle* de paramètre $\lambda > 0$:

$$f_Z(z) = \begin{cases} \lambda \exp(-z\lambda) & \forall z \in \mathbb{R}^+ \\ 0 & \text{sinon} \end{cases} \qquad (6.91)$$

Déterminons sa fonction de répartition. Par définition, $\forall z \in \mathbb{R}^+$:

$$F_Z(z) = \int_{-\infty}^{z} f_Z(u)\,du = \int_{0}^{z} f_Z(u)\,du \qquad (6.92)$$

Dès lors, il vient :

$$F_Z(z) = \int_{0}^{z} \lambda \exp(-u\lambda)\,du = \lambda \left[-\frac{1}{\lambda} \exp(-u\lambda) \right]_0^z = 1 - \exp(-z\lambda) \qquad (6.93)$$

La fonction de répartition de la variable Z est définie par :

$$F_Z(z) = 1 - \exp(-z\lambda) \quad \forall z \in \mathbb{R}^+ \qquad (6.94)$$

Dans les deux exemples précédents, nous sommes parvenus à obtenir une forme analytique (une « formule ») de la fonction de répartition en intégrant la fonction de densité. Mais ce n'est pas toujours le cas. Pour de nombreuses lois continues usuelles (loi normale, loi du khi-deux, loi de Student, loi de Fisher-Snedecor, etc.), il est impossible d'exprimer la primitive de la fonction de densité avec des fonctions simples (logarithme, puissance, exponentielle, etc.).

Remarque : Même s'il n'existe pas d'**expression analytique** pour la fonction de répartition, on peut toutefois l'approximer par des méthodes numériques. Pour toute valeur $x \in X(\Omega)$, on détermine alors numériquement la valeur de $F_X(x)$ sans avoir de formule générale pour cette fonction.

Exemple

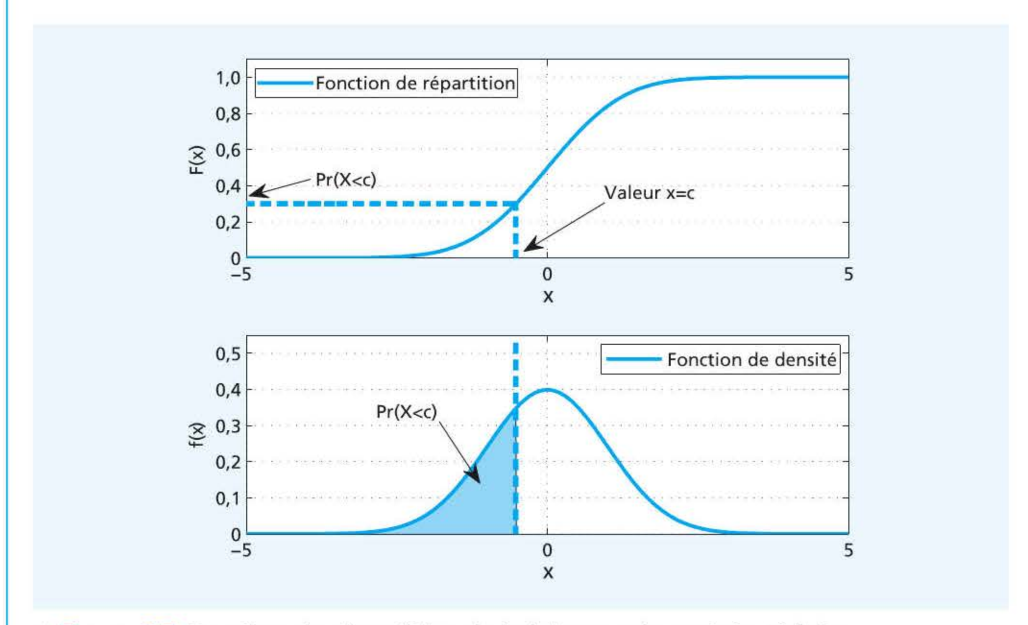
La fonction de densité, notée $\phi(x)$, d'une *loi normale centrée réduite*, notée $\mathcal{N}(0,1)$, est définie par :

$$\phi(x) = f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad \forall x \in \mathbb{R} \tag{6.95}$$

La fonction de répartition correspondante, notée $\Phi(x)$, n'a pas de forme analytique.

$$\Phi(x) = \int_{-\infty}^{x} \phi(u)\, du = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du = ? \tag{6.96}$$

Toutefois, il est possible de déterminer *numériquement* la valeur de $\Phi(x)$: pour cela on utilise soit des algorithmes d'approximation numérique d'intégrales (implémentés dans la plupart des logiciels mathématiques), soit des tables de loi (► chapitre 7). La figure 6.7 représente les fonctions de densité (graphique du haut) et de répartition (graphique du bas) de la loi *normale centrée réduite* obtenues à partir du logiciel Matlab. Ces fonctions sont représentées pour des valeurs de x allant de -5 à 5. Si l'on considère une valeur quelconque c sur cet intervalle, la valeur de la fonction de répartition $\Phi(c)$ correspond à la probabilité $\Pr(X \leq c)$. Par définition, cette probabilité est égale à l'aire sous la fonction de densité située à gauche de $x = c$.



▲ **Figure 6.7 Fonction de répartition de la loi normale centrée réduite**

Comme dans le cas de ces trois exemples, la fonction de répartition vérifie toujours les propriétés présentées dans la section 2.2 :

1. $F_X(x)$ est croissante avec x : $\partial F_X(x)/\partial x \geq 0$, $\forall x \in \mathbb{R}$.
2. $0 \leq F_X(x) \leq 1$, $\forall x \in \mathbb{R}$.
3. $\lim\limits_{x \to -\infty} F_X(x) = 0$.
4. $\lim\limits_{x \to +\infty} F_X(x) = 1$.

Le résultat selon lequel la fonction de répartition $F_X(x)$ est croissante avec x, tient au fait que sa dérivée première correspond à la densité (positive ou nulle par définition).

Propriété

Fonction de densité

Par construction, la fonction de densité correspond à la dérivée première de la fonction de répartition :

$$f_X(x) = \frac{\partial F_X(x)}{\partial x} \quad \forall x \in \mathbb{R} \tag{6.97}$$

Tout comme dans le cas des variables discrètes, il est possible « d'inverser » la fonction de répartition afin de déterminer la valeur de x qui correspond à une certaine probabilité cumulée $\alpha = \Pr(X \leq x)$, avec $\alpha \in [0,1]$. On obtient alors la fonction de répartition inverse ou le quantile d'ordre α. La définition du quantile est légèrement différente de celle présentée dans le cadre des variables discrètes (► section 2.2).

Définition 6.22

Si X est une variable aléatoire réelle, le **quantile d'ordre** α de sa loi de probabilité, noté $F_X^{-1}(\alpha)$, est la réalisation appartenant à $X(\Omega) \subseteq \mathbb{R}$ correspondant à une probabilité cumulée égale à α :

$$\Pr\left(X \leq F_X^{-1}(\alpha)\right) = F_X\left(F_X^{-1}(\alpha)\right) = \alpha \quad \forall \alpha \in [0,1] \tag{6.98}$$

Tout comme pour le cas des variables discrètes, il convient de noter qu'un quantile est une réalisation. Par conséquent, un quantile appartient au support $X(\Omega)$ et il n'est pas nécessairement défini sur $\mathbb{R}$ si $X(\Omega)$ est une partie de $\mathbb{R}$. Par exemple, dans le cas d'une loi uniforme continue sur le support [2,3] le quantile d'ordre $\alpha = 0$ est égal à 2 et le quantile d'ordre $\alpha = 1$ est égal à 3. Pour cette loi, tous les quantiles d'ordre $\alpha \in [0,1]$ appartiennent au support [2,3].

Exemple

On considère une variable aléatoire réelle Z à valeurs sur $Z(\Omega) = \mathbb{R}^+$ et distribuée selon une *loi exponentielle* de paramètre $\lambda = 2$, admettant une fonction de répartition définie par :

$$F_Z(z) = 1 - \exp(-z\lambda) \quad \forall z \in \mathbb{R}^+ \tag{6.99}$$

Posons $\alpha = F_X(z)$ et inversons la fonction $F_X(z)$. Il vient :

$$F_Z^{-1}(\alpha) = z = -\frac{\ln(1-\alpha)}{\lambda} \quad \forall \alpha \in [0,1] \tag{6.100}$$

Notons que $\forall \alpha \in [0,1]$, $F_Z^{-1}(\alpha) \in Z(\Omega) = \mathbb{R}^+$. Par exemple, le quantile d'ordre $\alpha = 5\ \%$ est égal à 0,0256 puisque :

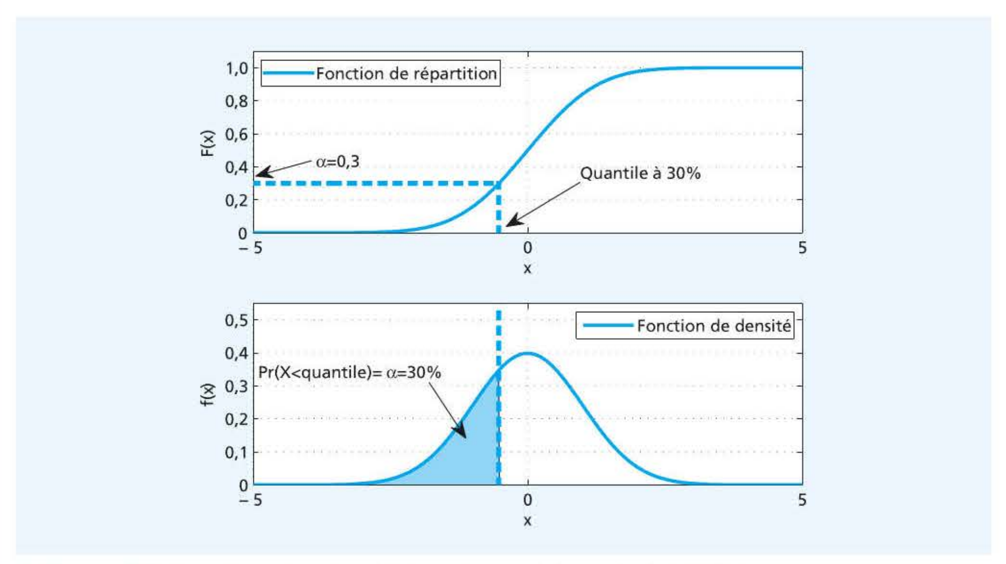
$$F_Z^{-1}(0{,}05) = -\frac{\ln(1-0.05)}{2} = 0{,}0256 \tag{6.101}$$

Le quantile $F_Z^{-1}(0{,}05)$ s'interprète de la façon suivante : il y a 5 % de chances que les réalisations de la variable aléatoire Z soient inférieures ou égales au seuil $F_Z^{-1}(0{,}05) = 0{,}0256$, *i.e.* $\Pr(Z \leq 0{,}0256) = 5\ \%$.

Remarque : Pour toutes les lois pour lesquelles il n'existe pas d'expression analytique de la fonction de répartition, il n'existe pas non plus d'expression analytique des quantiles. Ceux-ci sont alors approximés par des méthodes numériques.

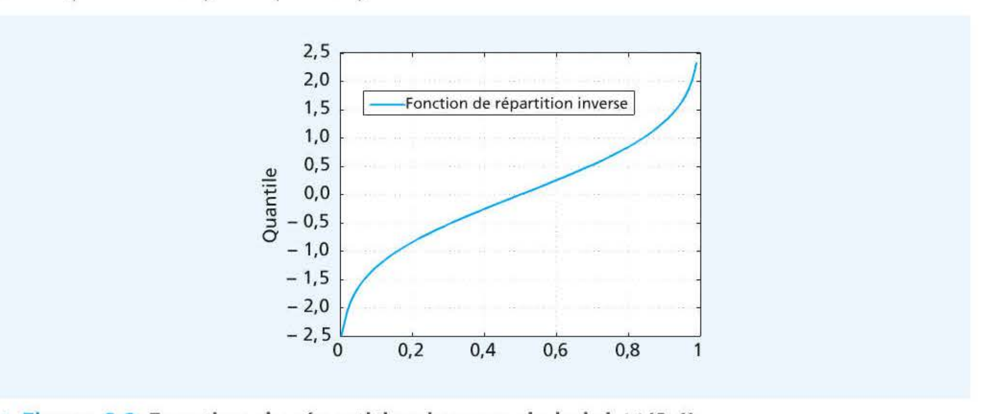
Exemple

On considère une variable aléatoire réelle $Y \in \mathbb{R}$ distribuée selon une *loi normale centrée réduite* $\mathcal{N}(0,1)$. Sa fonction de densité, notée $\phi(x)$, et sa fonction de répartition, notée $\Phi(x)$, sont représentées sur la figure 6.8. Rappelons que la fonction de densité est définie par $\phi(x) = (1/2\pi)^{-1/2} \exp\left(-x^2/2\right)$, mais que la fonction de répartition n'a pas d'expression analytique. Par définition, le quantile à $\alpha = 30\ \%$ peut être obtenu à partir de la fonction de répartition (graphique du haut de la figure 6.8). C'est la valeur de x telle que $\Phi(x) = 0{,}30$. En utilisant une table de loi ou un logiciel de statistique, on peut montrer que $\Phi^{-1}(0{,}30) = -0{,}5244$. Ce quantile peut aussi être obtenu à partir de la fonction de densité (graphique du bas de la figure 6.8). On cherche alors la valeur de x (axe des abscisses) telle que l'aire sous la fonction de densité située à gauche de x soit précisément égale à 30 %. Rappelons que cette aire représente la probabilité $\Pr(X \leq x) = \Phi(x)$. On retrouve naturellement la même valeur du quantile, à savoir $\Phi^{-1}(0{,}30) = -0{,}5244$.



▲ **Figure 6.8 Fonction de répartition et quantile de la loi $\mathcal{N}(0,1)$**

On peut bien évidemment reproduire ce raisonnement pour toutes les valeurs α comprises entre 0 et 1. On obtient alors la fonction de répartition inverse $\Phi^{-1}(\alpha)$. Celle de la loi normale centrée réduite est reproduite sur la figure 6.9. Pour chaque valeur $\alpha \in [0{,}1]$ sur l'axe des abscisses, cette fonction renvoie sur l'axe des ordonnées la valeur du quantile d'ordre α tel que $\Pr\left(X \leq \Phi^{-1}(\alpha)\right) = \Phi\left(\Phi^{-1}(\alpha)\right) = \alpha$.



▲ **Figure 6.9 Fonction de répartition inverse de la loi $\mathcal{N}(0,1)$**

3.3 Moments d'une variable aléatoire continue

La définition générale des **moments ordinaires et centrés** – intégrale de Riemann–Stieltjes, équations (6.29) et (6.30) – est valable quel que soit le type de variable aléatoire (discrète ou continue). Mais dans le cas des variables continues, nous pouvons également définir ces moments à partir de la fonction de densité.

Définition 6.23

Soit X une variable aléatoire réelle définie sur un support $X(\Omega) \subseteq \mathbb{R}$ et caractérisée par une fonction de densité $f_X(x)$. Le **moment ordinaire** (non centré) d'ordre $k \in \mathbb{N}$ de la loi de probabilité de X est défini par :

$$m_k = \mathbb{E}\left(X^k\right) = \int_{-\infty}^{+\infty} x^k \, f_X(x)\, dx \tag{6.102}$$

Le **moment centré d'ordre** $k \in \mathbb{N}$ de la loi de probabilité de X est défini par :

$$\mu_k = \mathbb{E}\left((X - \mathbb{E}(X))^k\right) = \int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^k \, f_X(x)\, dx \tag{6.103}$$

Exemple

On considère une variable aléatoire réelle Z distribuée selon une *loi exponentielle* de paramètre $\lambda > 0$ sur $Z(\Omega) = \mathbb{R}^+$ telle que :

$$f_Z(z) = \begin{cases} \lambda \exp(-z\lambda) & \forall z \in \mathbb{R}^+ \\ 0 & \text{sinon} \end{cases} \tag{6.104}$$

Déterminons ses moments ordinaires $m_1 = \mathbb{E}(Z)$ et $m_2 = \mathbb{E}(Z^2)$.

$$m_1 = \mathbb{E}(Z) = \int_{-\infty}^{+\infty} z\, f_Z(z)\, dz = \int_0^{+\infty} z\lambda \exp(-z\lambda)\, dz \tag{6.105}$$

En intégrant par parties avec $u = \lambda z$ et $v' = \exp(-z\lambda)$, il vient :

$$\mathbb{E}(Z) = [-z\exp(-\lambda z)]_0^{+\infty} + \int_0^{+\infty} \exp(-z\lambda)\, dz = 0 + \frac{1}{\lambda} = \frac{1}{\lambda} \tag{6.106}$$

puisque par définition de la densité, $\int_0^{+\infty} \exp(-z\lambda)\, dz = \lambda^{-1} \int_0^{+\infty} f_Z(z)\, dz = \lambda^{-1}$. De la même façon, déterminons $\mathbb{E}(Z^2)$.

$$m_2 = \mathbb{E}\left(Z^2\right) = \int_{-\infty}^{+\infty} z^2 \, f_Z(z)\, dz = \int_0^{+\infty} z^2 \lambda \exp(-z\lambda)\, dz \tag{6.107}$$

En intégrant par parties, on obtient :

$$\mathbb{E}(Z) = \left[-z^2 \exp(-\lambda z)\right]_0^{+\infty} + \int_0^{+\infty} 2z \, \exp(-z\lambda)\, dz = 0 + \frac{2}{\lambda^2} = \frac{2}{\lambda^2} \tag{6.108}$$

puisque $\int_0^{+\infty} z \, \exp(-z\lambda)\, dz = \lambda^{-1} \int_0^{+\infty} z\, f_Z(z)\, dz = \lambda^{-1} m_1 = \lambda^{-2}$.

Comme pour le cas des variables discrètes (▸ section 2.3), les moments ordinaires peuvent être obtenus à partir de la fonction génératrice des moments.

Définition 6.24

La **fonction génératrice des moments** d'une variable aléatoire réelle X est définie par :

$$M_X(t) = \mathbb{E}(\exp(tX)) = \int_{-\infty}^{+\infty} \exp(tx)\ f_X(x)\, dx \qquad \forall t \in \mathbb{R} \tag{6.109}$$

Rappelons que si le moment ordinaire d'ordre k existe, il correspond à la dérivée $k^{ème}$ de la fonction génératrice des moments évaluée au point $t = 0$.

$$m_k = \left. \frac{\partial^k M_X(t)}{\partial t^k} \right|_{t=0} \tag{6.110}$$

Exemple

On considère une variable aléatoire réelle Z distribuée selon une loi exponentielle de paramètre $\lambda > 0$ sur $Z(\Omega) = \mathbb{R}^+$. On admet que sa fonction génératrice des moments est définie par :

$$M_Z(t) = \left(1 - \frac{t}{\lambda}\right)^{-1} \tag{6.111}$$

Déterminons les moments ordinaires d'ordres 1 et 2 associés à la loi de probabilité de Z. Commençons par dériver la fonction génératrice des moments :

$$\frac{\partial M_Z(t)}{\partial t} = \frac{1}{\lambda}\left(1 - \frac{t}{\lambda}\right)^{-2} \qquad \frac{\partial^2 M_Z(t)}{\partial t^2} = \frac{2}{\lambda^2}\left(1 - \frac{t}{\lambda}\right)^{-2} \tag{6.112}$$

Dès lors, il vient :

$$m_1 = \mathbb{E}(Z) = \left. \frac{\partial M_X(t)}{\partial t} \right|_{t=0} = \frac{1}{\lambda}\left(1 - \frac{0}{\lambda}\right)^{-2} = \frac{1}{\lambda} \tag{6.113}$$

$$m_2 = \mathbb{E}\left(Z^2\right) = \left. \frac{\partial^2 M_X(t)}{\partial t^2} \right|_{t=0} = \frac{2}{\lambda^2}\left(1 - \frac{0}{\lambda}\right)^{-2} = \frac{2}{\lambda^2} \tag{6.114}$$

On vérifie que l'on retrouve les expressions de m_1 et m_2 obtenues précédemment.

3.4 Moments remarquables

Comme nous l'avons vu dans la section 2.3 pour le cas des variables discrètes, certains moments méritent généralement une attention toute particulière. C'est typiquement le cas de l'espérance m_1 et de la variance μ_2.

Définition 6.25

Soit X une variable aléatoire réelle continue, son **espérance** et sa **variance** sont définies par :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx \tag{6.115}$$

$$\mathbb{V}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^2\ f_X(x)\, dx \tag{6.116}$$

L'interprétation et les propriétés de ces deux moments sont identiques à celles que nous avions présentées dans la section 2.3 pour le cas des variables discrètes. Nous

n'y reviendrons pas. Toutefois, dans le cas des variables continues en plus de ces deux moments, on s'intéresse parfois (par exemple en finance ou en gestion des risques) à des transformées des moments centrés d'ordre 3 et 4, à savoir la **skewness** (ou coefficient de dissymétrie) et la **kurtosis** (ou coefficient d'aplatissement). Il convient une nouvelle fois de bien distinguer les concepts de skewness et de kurtosis associés à une variable aléatoire, et les concepts de skewness et de kurtosis empiriques associés à un échantillon ou une population (▶ chapitre 1).

3.4.1 Skewness

La skewness est un indicateur de la dissymétrie de la distribution.

Définition 6.26

La **skewness** (ou coefficient de dissymétrie) d'une variable aléatoire X est définie par :

$$\text{skewness} = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mathbb{E}((X - \mathbb{E}(X))^3)}{\mathbb{V}(X)^{3/2}} \tag{6.117}$$

Notons que cette définition est valable quel que soit le type (discret ou continu) de variable aléatoire considérée[11]. Une autre façon d'interpréter la skewness consiste à remarquer que si l'on pose $\mathbb{E}(X) = m_1$ et $\mathbb{V}(X) = \mu_2$, alors il vient :

$$\text{skewness} = \mathbb{E}\left(\left(\frac{X - m_1}{\mu_2^{1/2}}\right)^3\right) \tag{6.118}$$

La skewness correspond donc au moment centré d'ordre 3 de la variable centrée réduite $(X - m_1)/\mu_2^{1/2}$.

Propriété

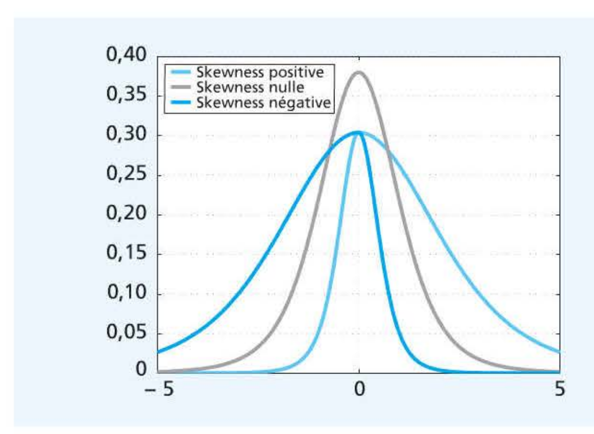
Skewness

La skewness est un indicateur de la symétrie de la fonction de densité par rapport à $\mathbb{E}(X) = m_1$. En effet, si la fonction de densité est symétrique, c'est-à-dire si $f_X(m_1 - x) = f_X(m_1 + x)\ \forall x \in \mathbb{R}$, alors la skewness est nulle[12].

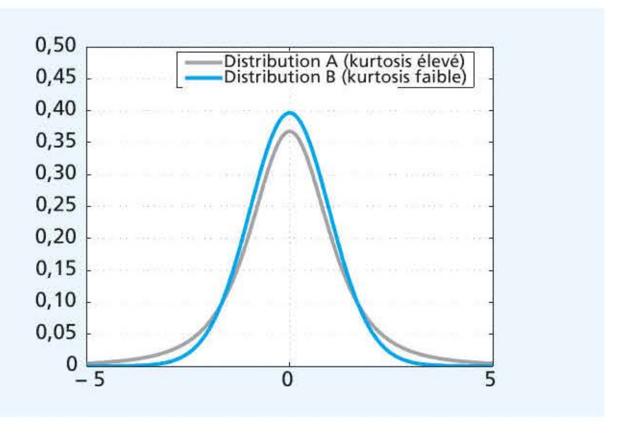
Dit autrement, une « skewness nulle » indique que l'on a autant de chances d'obtenir des réalisations inférieures à l'espérance $\mathbb{E}(X)$ que d'obtenir des réalisations supérieures à l'espérance, *i.e.* $\Pr(X \leq \mathbb{E}(X)) = \Pr(X \geq \mathbb{E}(X)) = 1/2$. Si la skewness est *positive*, la queue de distribution est étalée vers la droite, comme l'illustre la figure 6.10. Cela signifie que la probabilité d'obtenir des réalisations supérieures à l'espérance $\mathbb{E}(X)$ est supérieure à la probabilité d'obtenir des réalisations inférieures à $\mathbb{E}(X)$, *i.e.* $\Pr(X \geq \mathbb{E}(X)) > \Pr(X \leq \mathbb{E}(X))$.

11 La skewness et la kurtosis peuvent être calculées pour des lois de probabilité discrètes. Toutefois, dans la pratique, on s'intéresse plus souvent aux phénomènes de dissymétrie ou d'applatissement dans le cadre de distributions continues.

12 Puisque la variance est toujours positive, la skewness est nulle dès lors que le moment ordinaire d'ordre 3 est nul, *i.e.* $\mu_3 = \mathbb{E}((X - \mathbb{E}(X))^3) = 0$.



▲ **Figure 6.10** Skewness et dissymétrie de la fonction de densité



▲ **Figure 6.11** Kurtosis et aplatissement des queues de distribution

3.4.2 Kurtosis

La kurtosis est un indicateur de l'aplatissement des queues de la distribution.

Définition 6.27

La **kurtosis** (ou coefficient d'aplatissement) d'une variable aléatoire X est définie par :

$$\text{kurtosis} = \frac{\mu_4}{\mu_2^2} = \frac{\mathbb{E}((X - \mathbb{E}(X))^4)}{\mathbb{V}(X)^2} \tag{6.119}$$

On peut montrer que la kurtosis correspond au moment centré d'ordre 4 de la variable centrée réduite $(X - m_1)/\mu_2^{1/2}$:

$$\text{kurtosis} = \mathbb{E}\left(\left(\frac{X - m_1}{\mu_2^{1/2}}\right)^4\right) \tag{6.120}$$

Propriété

La kurtosis est un indicateur de l'aplatissement des queues de la distribution : plus la kurtosis est élevé, plus la probabilité d'apparition d'événements « extrêmes » (réalisations très grandes en valeur absolue) est élevée.

Soient deux distributions symétriques A et B de même espérance telles que le kurtosis de A est supérieur au kurtosis de B. Comme l'illustre la figure 6.11, les queues gauche et droite[13] de la distribution A sont plus « épaisses » que celles de la distribution B. Quelle est l'implication de ce résultat ? Considérons la probabilité d'obtenir des réalisations positives très élevées, par exemple supérieures à 3 dans ce cas. L'aire sous la densité de A (courbe rouge) située à droite de la valeur 3 est plus grande que l'aire sous la densité de B (courbe bleue). Par conséquent $\Pr(A \geq 3) > \Pr(B \geq 3)$: la probabilité d'apparition de très fortes réalisations positives est plus élevée avec la distribution A qu'avec la distribution B. On peut faire le même raisonnement pour des réalisations fortement négatives.

13 Dans le cas de distributions dissymétriques, on aurait pu avoir une configuration où seule la queue gauche (ou droite) de la distribution de A est plus « épaisse » que celle de B.

On compare généralement la kurtosis d'une distribution à cellle de la loi normale. Pour une loi normale, la kurtosis est égale à 3. Suivant la valeur de la kurtosis, on distingue trois types de distributions (▸ chapitre 1) :

- Si la kurtosis est supérieure à 3, on dit que la distribution est **leptokurtique**.
- Si la kurtosis est égale à 3, on dit que la distribution est **mésokurtique**.
- Si la kurtosis est inférieure à 3, on dit que la distribution est **platykurtique**.

Si une distribution est leptokurtique, cela signifie que sa kurtosis est supérieure à celle de la loi normale. Par conséquent, la probabilité d'apparition d'événements « extrêmes » est plus élevée avec cette distribution qu'avec une distribution normale. En finance, on observe généralement que c'est typiquement le cas pour la distribution des rendements de la plupart des actifs financiers. Si une distribution est mésokurtique, cela signifie au contraire que sa kurtosis est égale à celle de la loi normale. On peut aussi définir une kurtosis normalisée égale à $\mu_4/\mu_2^2 - 3$. La caractérisation de la distribution se fait alors par comparaison de la kurtosis normalisée par rapport à 0.

4 Comparaison des variables continues et discrètes

L'objectif de cette section est de proposer une comparaison synthétique des variables aléatoires discrètes et des variables aléatoires continues.

Propriété

Une loi de probabilité discrète ou continue peut être caractérisée de façon équivalente par (i) sa fonction de masse ou de densité, (ii) sa fonction de répartition ou (iii) la population de ses moments.

Le tableau 6.5 résume les principales différences entre les variables aléatoires continues et les variables aléatoires discrètes. Afin de simplifier les notations pour les variables aléatoires discrètes nous ne présenterons les principales formules que dans le cas où le support $X(\Omega)$ est fini dénombrable de dimension n.

▼ **Tableau 6.5** Principales propriétés des variables aléatoires continues et discrètes

Variable aléatoire discrète	Variable aléatoire continue
Support	
Fini (ou infini dénombrable)	Infini non dénombrable
$X(\Omega) = \{x_1,...,x_n\}$	$X(\Omega) \subseteq \mathbb{R}$
Loi de probabilité	
Fonction de masse	Fonction de densité
$f_X(x) = \Pr(X = x)$	$f_X(x)$
$0 \leq f_X(x) \leq 1 \ \forall x \in X(\Omega)$	$f_X(x) \geq 0 \ \forall x \in X(\Omega)$
$\sum_{i=1}^{n} f_X(x_i) = 1$	$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$

Variable aléatoire discrète	Variable aléatoire continue
Fonction de répartition	
$F_X(x) = \Pr(X \leq x)$	
$0 \leq F_X(x) \leq 1 \quad \forall x \in \mathbb{R}$	
$F_X(x) = \sum_{x_i \leq x} \Pr(X = x_i)$	$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du$
Quantile d'ordre α ou fonction de répartition inverse	
$\alpha \in [0,1] \quad F_X^{-1}(\alpha) \in X(\Omega)$	
Plus petite réalisation $F_X^{-1}(\alpha)$ telle que $\Pr\left(X \leq F_X^{-1}(\alpha)\right) \geq \alpha$	Réalisation $F_X^{-1}(\alpha)$ telle que $\Pr\left(X \leq F_X^{-1}(\alpha)\right) = \alpha$

Le tableau 6.6 reprend les différentes notions relatives aux moments (ordinaires et centrés), ainsi que les définitions et les propriétés de l'espérance et de la variance, pour les variables discrètes et continues.

▼ **Tableau 6.6 Moments des variables aléatoires continues et discrètes**

Variable aléatoire discrète	Variable aléatoire continue
Moments ordinaires (non centrés)	
$m_k = \mathbb{E}\left(X^k\right) \quad \forall k \in \mathbb{N}$	
$m_k = \sum_{i=1}^{n} x_i^k \Pr(X = x_i)$	$m_k = \int_{-\infty}^{+\infty} x^k f_X(x)\, dx$
Moments centrés	
$\mu_k = \mathbb{E}((X - \mathbb{E}(X))^k) \quad \forall k \in \mathbb{N}$	
$\mu_k = \sum_{i=1}^{n} (x_i - \mathbb{E}(X))^k \Pr(X = x_i)$	$\mu_k = \int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^k f_X(x)\, dx$
Fonction génératrice de moments	
$M_X(t) = \mathbb{E}(\exp(tX)) \quad \forall t \in \mathbb{R}$	
$M_X(t) = \sum_{i=1}^{n} \exp(tx_i) \Pr(X = x_i)$	$M_X(t) = \int_{-\infty}^{+\infty} \exp(tx)\, f_X(x)\, dx$
Espérance $\mathbb{E}(X) = m_1$	
$\mathbb{E}(X) = \sum_{i=1}^{n} x_i \Pr(X = x_i)$	$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f_X(x)\, dx$
$\mathbb{E}(g(X)) = \sum_{i=1}^{n} g(x_i) \Pr(X = x_i)$	$\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x) f_X(x)\, dx$
$\forall (a,b) \in \mathbb{R}^2, \ \mathbb{E}(a + bX) = a + b\mathbb{E}(X)$	
Variance $\mathbb{V}(X) = \mu_2$	
$\mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}\left(X^2\right) - (\mathbb{E}(X))^2$	
$\mathbb{V}(X) = \sum_{i=1}^{n} (x_i - \mathbb{E}(X))^2 \Pr(X = x_i)$	$\mathbb{V}(X) = \int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^2 f_X(x)\, dx$
$\forall (a,b) \in \mathbb{R}^2, \ \mathbb{V}(a + bX) = b^2 \mathbb{V}(X)$	

5 Couples et vecteurs de variables aléatoires

Considérons à présent deux variables aléatoires (discrètes ou continues) X et Y respectivement définies sur $X(\Omega)$ et $Y(\Omega)$. On peut alors définir le **couple** de variables aléatoires (X,Y) de la façon suivante.

Définition 6.28

Les réalisations du **couple de variables aléatoires** (X,Y) appartiennent à l'univers des réalisations $X(\Omega) \times Y(\Omega)$.

Le symbole $\times$ correspond au *produit cartesien*. Cela signifie que l'univers des réalisations $X(\Omega) \times Y(\Omega)$ (prononcer $X(\Omega)$ *croix* $Y(\Omega)$) correspond à l'ensemble de tous les couples de réalisations (x,y) où $x \in X(\Omega)$ et $y \in Y(\Omega)$. À partir d'un couple de variables aléatoires, il est possible de définir trois notions de distribution :

- la distribution jointe ;
- la distribution marginale ;
- la distribution conditionnelle.

5.1 Loi jointe et loi marginale

Nous allons distinguer les couples de variables aléatoires discrètes des couples de variables aléatoires continues.

5.1.1 Cas d'un couple de variables discrètes

La loi de probabilité jointe (ou distribution jointe) d'un couple de variables aléatoires discrètes est définie de la façon suivante.

Définition 6.29

L'application $\Pr((X = x_i) \cap (Y = y_i))$, $\forall (x_i,y_i) \in X(\Omega) \times Y(\Omega)$ définit la **loi de probabilité jointe** du couple de variables aléatoires discrètes (X,Y). Puisque les réalisations forment un système complet :

$$\sum_{(x_i,y_i)\in X(\Omega)\times Y(\Omega)} \Pr((X = x_i) \cap (Y = y_i)) = 1 \qquad (6.121)$$

La quantité $\Pr((X = x_i) \cap (Y = y_i))$ correspond à la **probabilité jointe** d'observer à la fois $X = x_i$ et $Y = y_i$. On peut la noter de différentes façons :

$$\Pr(X = x_i, Y = y_i) \equiv \Pr(x_i,y_i) \equiv \Pr((X = x_i) \cap (Y = y_i)) \qquad (6.122)$$

Remarque : Une autre façon de se représenter un couple de variables aléatoires consiste à supposer que le couple (X,Y) est un vecteur de variables aléatoires $Z = (X,Y)^\top$ de dimension 2×1 défini sur l'univers des réalisations $Z(\Omega) = X(\Omega) \times Y(\Omega)$. Le symbole $\top$ correspond à la transposée. Pour chaque réalisation $z_i = (x_i, y_i)^\top \in Z(\Omega)$, on associe une probabilité $\Pr(Z = z_i)$. L'ensemble des probabilités $\Pr(Z = z_i)$, $\forall z_i \in Z(\Omega)$, permet de caractériser la loi de probabilité (ou distribution) du vecteur aléatoire Z. Cette loi de probabilité correspond à la loi de probabilité jointe du couple (X,Y).

La loi de probabilité jointe peut être représentée de façon équivalente par (1) les probabilités jointes, (2) la fonction de répartition associée, ou (3) la population des moments associés. Le principe de construction de la fonction de répartition jointe et des moments associés est similaire à celui présenté dans le cadre univarié (▶ section 2.2).

Pour un couple (ou un vecteur) de variables aléatoires (X,Y), nous pouvons définir les lois de probabilité marginales des variables aléatoires X et Y. Ces lois correspondent aux lois des variables X et Y considérées en isolation.

Définition 6.30

Soit (X,Y) un couple de variables aléatoires défini sur le support $X(\Omega) \times Y(\Omega)$. On appelle **lois de probabilité marginales** de X et de Y, les applications respectivement définies par :

$$\Pr(X = x_i) = \sum_{y_i \in Y(\Omega)} \Pr((X = x_i) \cap (Y = y_i)) \tag{6.123}$$

$$\Pr(Y = y_i) = \sum_{x_i \in X(\Omega)} \Pr((X = x_i) \cap (Y = y_i)) \tag{6.124}$$

La **probabilité marginale** $\Pr(X = x_i)$ correspond ainsi à la somme des probabilités jointes d'observer $X = x_i$ conjointement à toutes les réalisations possibles de Y *i.e.* $Y = y_1, \ldots, Y = y_n$ si $Y(\Omega) = \{y_1, \ldots, y_n\}$.

Par construction, la somme des probabilités marginales $\Pr(X = x_i)$ associées à toutes les réalisations $x_i \in X(\Omega)$ est égale à 1 :

$$\sum_{x_i \in X(\Omega)} \Pr(X = x_i) = \sum_{x_i \in X(\Omega)} \sum_{y_i \in Y(\Omega)} \Pr((X = x_i) \cap (Y = y_i)) = 1 \tag{6.125}$$

De la même façon :

$$\sum_{y_i \in Y(\Omega)} \Pr(Y = y_i) = \sum_{y_i \in Y(\Omega)} \sum_{x_i \in X(\Omega)} \Pr((X = x_i) \cap (Y = y_i)) = 1 \tag{6.126}$$

Remarque : Les lois de *probabilité marginales* des variables aléatoires X et Y, décrites par les équations (6.123) et (6.124) à partir de la loi jointe du couple (X,Y), correspondent aux lois de probabilité des variables X et Y *considérées en isolation*.

Exemple

On considère deux variables aléatoires indépendantes X et Y respectivement définies sur $X(\Omega) = \{a,b\}$ et $Y(\Omega) = \{1,2\}$, telles que :

$$\Pr(X = a) = 0{,}2 \qquad \Pr(X = b) = 0{,}8 \tag{6.127}$$

$$\Pr(Y = 1) = 0{,}7 \qquad \Pr(Y = 2) = 0{,}3 \tag{6.128}$$

On admet que la loi de probabilité jointe du couple (X,Y), définie sur $X(\Omega) \times Y(\Omega) = \{\{a,1\},\{a,2\},\{b,1\},\{b,2\}\}$, est définie par :

$$\Pr((X = a) \cap (Y = 1)) = 0{,}14 \qquad \Pr((X = a) \cap (Y = 2)) = 0{,}06 \tag{6.129}$$

$$\Pr((X = b) \cap (Y = 1)) = 0{,}56 \qquad \Pr((X = b) \cap (Y = 2)) = 0{,}24 \tag{6.130}$$

La probabilité marginale d'observer $X = a$ est égale à :

$$\begin{aligned}\Pr(X = a) &= \Pr((X = a) \cap (Y = 1)) + \Pr((X = a) \cap (Y = 2)) \\ &= 0{,}14 + 0{,}06 = 0{,}2\end{aligned} \tag{6.131}$$

On vérifie que l'on retrouve la même valeur que dans l'équation (6.127).

Les notions de distribution jointe et de distribution marginale peuvent être étendues à un **vecteur** de $k \geq 2$ variables aléatoires discrètes.

Définition 6.31

Soit $x_i = (x_{1,i},\ldots,x_{k,i})^\top$ un vecteur de réalisations du **vecteur** de variables aléatoires discrètes $X = (X_1,\ldots,X_k)^\top$ défini sur le support $X(\Omega) = X_1(\Omega) \times \ldots \times X_k(\Omega)$, alors la *loi de probabilité jointe* des éléments du vecteur X est définie par l'application :

$$\Pr(X = x_i) = \Pr((X_1 = x_{1,i}) \cap \ldots \cap (X_k = x_{k,i})) \tag{6.132}$$

La probabilité marginale pour toute variable X_j est définie par l'application :

$$\Pr\left(X_j = x_{j,i}\right) = \underbrace{\sum_{x_{1,i} \in X_1(\Omega)} \ldots \sum_{x_{k,i} \in X_k(\Omega)}}_{k-1 \text{ sommes (on exclut } X_j)} \Pr\left(\left(X_j = x_{j,i}\right) \cap \underbrace{(X_1 = x_{1,i}) \cap \ldots \cap (X_k = x_{k,i})}_{k-1 \text{ événements (on exclut } X_j)}\right) \tag{6.133}$$

Par exemple, si $k = 3$ et $X_1(\Omega) = X_2(\Omega) = X_3(\Omega) = \{1,..,n\}$, alors les probabilités marginales des variables X_1 et X_2 sont définies par :

$$\Pr(X_1 = i) = \sum_{j=1}^{n} \sum_{z=1}^{n} \Pr((X_1 = i) \cap (X_2 = j) \cap (X_3 = z)) \quad \forall i = 1,\ldots,n \tag{6.134}$$

$$\Pr(X_2 = j) = \sum_{i=1}^{n} \sum_{z=1}^{n} \Pr((X_2 = j) \cap (X_1 = i) \cap (X_3 = z)) \quad \forall j = 1,\ldots,n \tag{6.135}$$

5.1.2 Cas d'un couple de variables continues

Dans le cas d'un couple de variables aléatoires continues, la logique est la même. On peut définir la densité jointe du couple (X,Y) comme suit.

Définition 6.32

Soit (X,Y) un couple de variables aléatoires réelles continues définies sur le support $X(\Omega) \times Y(\Omega) \subseteq \mathbb{R}^2$. La **loi jointe** du couple (X,Y) admet une **fonction de densité jointe**, notée $f_{X,Y}(x,y)$, si cette fonction est définie sur $X(\Omega) \times Y(\Omega)$, positive ou nulle, intégrable et telle que $\forall (a_x,b_x) \in X(\Omega)^2$ et $\forall \left(a_y,b_y\right) \in Y(\Omega)^2$:

$$\Pr\left((a_x \leq X \leq b_x) \cap \left(a_y \leq Y \leq b_y\right)\right) = \int_{a_x}^{b_x} \int_{a_y}^{b_y} f_{X,Y}(x,y)\, dxdy \qquad (6.136)$$

La fonction de densité jointe possède les mêmes propriétés qu'une fonction de densité. Elle est toujours positive ou nulle sur le support du couple (X,Y) et vérifie :

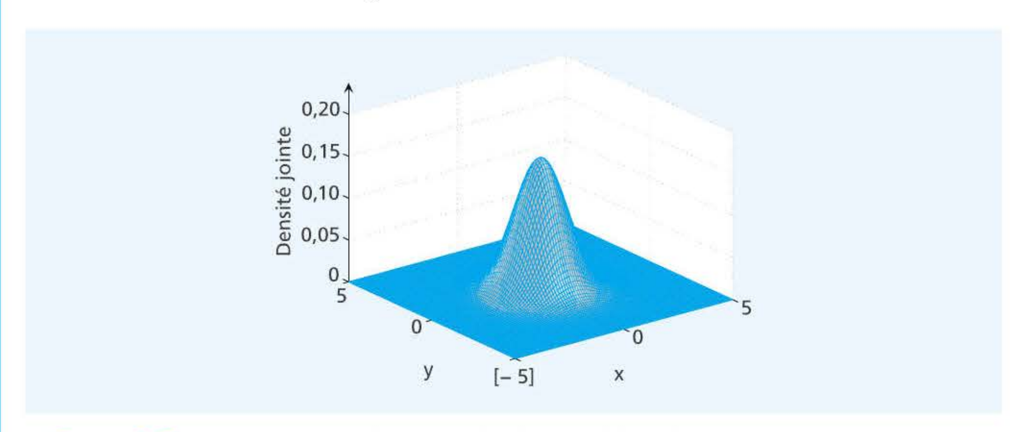
$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\, dxdy = 1 \qquad (6.137)$$

Exemple

On considère un couple de variables aléatoires réelles continues (X,Y) définies sur $\mathbb{R}^2$, admettant une distribution jointe normale bivariée telle que :

$$f_{X,Y}(x,y) = \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right) \quad \forall (x,y) \in \mathbb{R}^2 \qquad (6.138)$$

Cette fonction de densité jointe est représentée sur la figure 6.12. Pour toute valeur de x (axe X) et toute valeur de y (axe Y), correspond une valeur de la densité jointe (axe vertical). Si l'on souhaite calculer la probabilité jointe d'observer $X \leq a$ et $Y \leq b$, il suffit d'évaluer le volume sous la fonction de densité jointe pour des valeurs de X inférieures ou égales à a et des valeurs de Y inférieures ou égales à b.



▲ **Figure 6.12 Densité jointe d'un couple de variables aléatoires normales**

Comme pour toute loi, on peut représenter de façon équivalente la loi jointe du couple (X,Y) par (i) sa fonction de densité, (ii) sa *fonction de répartition* ou (iii) la *population des moments* associés.

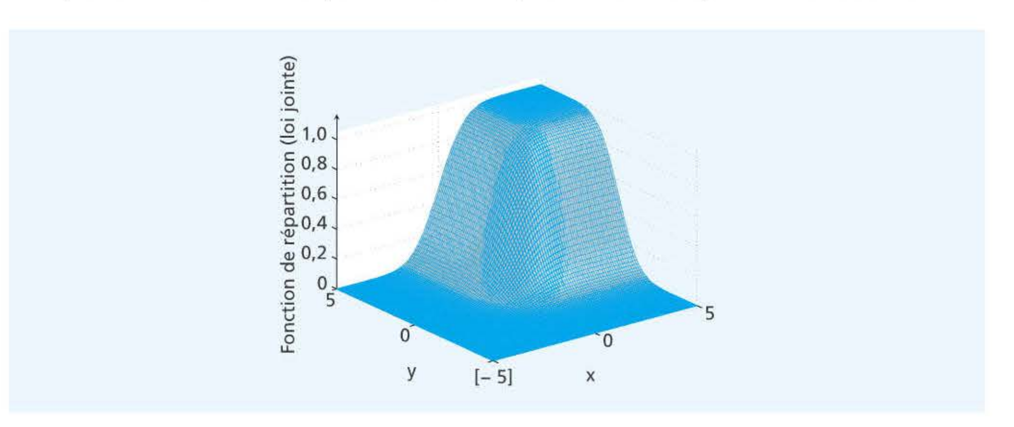
Définition 6.33

La **fonction de répartition de la loi jointe** du couple (X,Y), notée $F_{X,Y}(x,y)$ correspond à la probabilité que les variables X et Y soient conjointement inférieures où égales à $(x,y) \in \mathbb{R}^2$:

$$F_{X,Y}(x,y) = \Pr((X \leq x) \cap (Y \leq y)) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)\, du dv \qquad (6.139)$$

Exemple

On considère un couple de variables aléatoires réelles (X,Y) définies sur $\mathbb{R}^2$, admettant une distribution jointe normale bivariée centrée réduite. Il n'existe pas d'expression analytique de la fonction de répartition de cette loi. Toutefois, il est possible de l'approximer numériquement en utilisant des fonctions prédéfinies disponibles dans la plupart des logiciels d'économétrie et de mathématique. La figure 6.13 représente cette fonction de répartition. On vérifie que lorsque les valeurs de x augmentent, à y constant, la fonction de répartition augmente. Lorsque les deux valeurs x et y tendent vers $+\infty$, la fonction de répartition tend vers 1. Lorsque les deux valeurs x et y tendent vers $-\infty$, la fonction de répartition tend vers 0.



▲ **Figure 6.13 Fonction de répartition de la loi jointe d'un couple de variables aléatoires normales**

À partir de la densité jointe, on peut déterminer les *densités marginales* des variables X et Y.

Définition 6.34

Soit (X,Y) un couple de variables aléatoires réelles définies sur le support $X(\Omega) \times Y(\Omega) \subseteq \mathbb{R}^2$. Les **fonctions de densité marginales** des variables X et Y, notées $f_X(x)$ et $f_Y(y)$, sont définies par :

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\, dy \qquad f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\, dx \qquad (6.140)$$

Exemple

On considère un couple de variables aléatoires réelles continues (X,Y) définies sur $\mathbb{R}^2$, admettant une distribution jointe *normale bivariée* standard telle que :

$$f_{X,Y}(x,y) = \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right) \quad \forall (x,y) \in \mathbb{R}^2 \qquad (6.141)$$

Déterminons la densité marginale de la variable X. Par définition, il vient :

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy = \int_{-\infty}^{+\infty} \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right) dy \quad (6.142)$$

Sachant que $\exp(a+b) = \exp(a) \times \exp(b)$, on peut réécrire cette expression comme :

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy \quad (6.143)$$

Par définition, $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-y^2/2\right) dy = 1$. Par conséquent :

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad (6.144)$$

On retrouve l'expression de la densité d'une loi normale centrée réduite. La loi marginale de X correspond à la loi de cette variable prise en isolation.

Les notions de densité jointe et de densité marginale peuvent être étendues à un *vecteur* de $k \geq 2$ variables aléatoires continues.

Définition 6.35

Soit $X = (X_1, \ldots, X_k)^\top$ un **vecteur de variables aléatoires réelles** défini sur le support $X(\Omega) \subseteq \mathbb{R}^k$. La fonction de densité jointe, notée $f_{X_1,\ldots,X_k}(x_1, \ldots, x_k)$, est telle que :

$$\Pr((X_1 \leq x_1) \cap \ldots \cap (X_k \leq x_k)) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_k} f_{X_1,\ldots,X_k}(u_1,\ldots,u_k)\, du_1 \ldots du_k \quad (6.145)$$

La fonction de densité *marginale* associée à la variable X_j est définie par :

$$f_{X_j}(x) = \underbrace{\int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty}}_{k-1 \text{ intégrales}} f_{X_1,\ldots,X_k} \underbrace{(u_1,\ldots,x,\ldots,u_k)}_{x \text{ à la } j^{ème} \text{ position}} \quad \underbrace{du_1 \ldots du_k}_{k-1 \text{ termes (on exclut } du_j)} \quad (6.146)$$

Par exemple si $k = 3$, les densités marginales des variables X_1 et X_2 sont définies par :

$$f_{X_1}(x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X_1,X_2,X_3}(x,u_2,u_3)\, du_2 du_3 \quad \forall x \in X_1(\Omega) \quad (6.147)$$

$$f_{X_2}(x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X_1,X_2,X_3}(u_1,x,u_3)\, du_1 du_3 \quad \forall x \in X_2(\Omega) \quad (6.148)$$

5.2 Moments d'un vecteur de variables aléatoires

Les moments associés aux distributions marginales des variables X et Y permettent de décrire ces deux distributions. Mais quels moments utiliser pour décrire la distribution jointe du couple (X,Y) et plus précisément le lien entre ces deux variables ? On utilise pour cela des **moments croisés** basés sur le produit des variables aléatoires $X \times Y$. Le plus utilisé des moments croisés est la **covariance** (► chapitre 2).

5.2.1 Covariance

Définition 6.36

La **covariance** de deux variables aléatoires X et Y est définie par :

$$\mathbb{C}ov\,(X,Y) = \mathbb{E}\,[(X - \mathbb{E}\,(X)) \times (Y - \mathbb{E}\,(Y))] \tag{6.149}$$

ou de façon équivalente par :

$$\mathbb{C}ov\,(X,Y) = \mathbb{E}\,(XY) - \mathbb{E}\,(X)\,\mathbb{E}\,(Y) \tag{6.150}$$

La covariance n'est rien d'autre que l'espérance du produit des variables centrées sur leurs espérances respectives. Ce moment permet d'évaluer le sens de variation des variables X et Y.

Si la covariance est positive, cela traduit le fait que les réalisations des deux variables ont tendance à être simultanément au-dessus ou en dessous de leurs espérances respectives. Dit autrement, si la covariance est positive, les réalisations des variables X et Y évoluent « dans le même sens » : elles ont tendance à être élevées ou faibles en même temps.

Si la covariance est négative, les réalisations ont tendance à évoluer « *en sens opposé* » : lorsque les réalisations de X sont élevées par rapport à son espérance, celles de Y ont, au contraire, tendance à être plus faibles que son espérance. Enfin, si la covariance est nulle cela traduit l'indépendance des deux variables X et Y.

Remarque : Il est important de noter que *l'indépendance implique la nullité de la covariance*, mais que la *réciproque n'est pas nécessairement vraie*.

$$\text{si } X \text{ et } Y \text{ sont indépendantes alors } \mathbb{C}ov\,(X,Y) = 0$$

$$\text{si } \mathbb{C}ov\,(X,Y) = 0 \text{ alors } X \text{ et } Y \text{ ne sont pas nécessairement indépendantes}$$

Ce résultat s'explique par le fait que la covariance est une mesure particulière de la dépendance qui peut exister entre X et Y : c'est une mesure de la **dépendance linéaire** puisqu'elle est définie comme une espérance. Or, même s'il n'existe pas de dépendance linéaire entre X et Y, du type $Y = a + bX$, il peut tout à fait exister d'autres formes de dépendances non-linéaires, par exemple $Y = X^2$ ou $Y = \ln(X)$. Ainsi, la nullité de la covariance ne garantit pas l'absence de dépendance au sens large (et donc l'indépendance), mais uniquement l'absence de dépendance linéaire. Notons que la condition $\mathbb{C}ov\,(X,Y) = 0$ est équivalente à la condition $\mathbb{E}\,(XY) = \mathbb{E}\,(X)\,\mathbb{E}\,(Y)$.

Remarque : La covariance est une mesure symétrique $\mathbb{C}ov\,(X,Y) = \mathbb{C}ov\,(Y,X)$. Par définition $\mathbb{C}ov\,(X,X) = \mathbb{E}(\,(X - \mathbb{E}\,(X))^2) = \mathbb{V}\,(X)$.

Par définition $\mathbb{C}ov\,(X,Y) \in \mathbb{R}$. Une mesure normalisée sur $[-1,1]$ de dépendance linéaire est donnée par la **corrélation** (► chapitre 2).

Définition 6.37

La **corrélation** entre deux variables aléatoires X et Y est définie par :

$$corr\,(X,Y) = \frac{\mathbb{C}ov\,(X,Y)}{\sigma\,(X)\,\sigma\,(Y)} \tag{6.151}$$

où $\sigma\,(X) = \sqrt{\mathbb{V}\,(X)}$ et $\sigma\,(Y) = \sqrt{\mathbb{V}\,(Y)}$ désignent les écarts-types des variables X et Y. Par construction, $corr\,(X,Y) \in [-1,1]$.

Comment déterminer la covariance (et donc la corrélation) dans le cas de variables aléatoires discrètes et de variables aléatoires continues ? Pour cela, on utilise la distribution jointe du couple (X,Y).

Définition 6.38

La covariance de deux variables aléatoires *discrètes* X et Y est égale à :

$$\mathbb{C}ov\,(X,Y) = \sum_{x_i \in X(\Omega)} \sum_{y_i \in Y(\Omega)} x_i\, y_i \Pr\left((X = x_i) \cap (Y = y_i)\right) - \mathbb{E}\,(X) \times \mathbb{E}\,(Y) \quad (6.152)$$

La covariance de deux variables aléatoires *continues* X et Y est égale à :

$$\mathbb{C}ov\,(X,Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x\, y\, f_{X,Y}\,(x,y)\; dx\, dy - \mathbb{E}\,(X) \times \mathbb{E}\,(Y) \quad (6.153)$$

Exemple

On considère deux variables aléatoires indépendantes X et Y respectivement définies sur $X(\Omega) = \{10,20\}$ et $Y(\Omega) = \{1,2\}$ telles que :

$$\Pr(X = 10) = 0{,}2 \qquad \Pr(Y = 1) = 0{,}7 \quad (6.154)$$

On admet que la loi de probabilité jointe du couple (X,Y) sur $X(\Omega) \times Y(\Omega) = \{\{10,1\},\{10,2\},\{20,1\},\{20,2\}\}$, est définie par :

$$\Pr((X = 10) \cap (Y = 1)) = 0{,}14 \quad \Pr((X = 10) \cap (Y = 2)) = 0{,}06 \quad (6.155)$$

$$\Pr((X = 20) \cap (Y = 1)) = 0{,}56 \quad \Pr((X = 20) \cap (Y = 2)) = 0{,}24 \quad (6.156)$$

On montre que :

$$\mathbb{E}\,(X) = 10 \times \Pr(X = 10) + 20 \times \Pr(X = 20) = 18 \quad (6.157)$$

$$\mathbb{E}\,(Y) = 1 \times \Pr(Y = 1) + 2 \times \Pr(Y = 2) = 1{,}3 \quad (6.158)$$

La covariance entre X et Y est définie par :

$$\mathbb{C}ov\,(X,Y) = \sum_{i=1}^{4} \sum_{j=1}^{4} x_i\, y_i \Pr((X = x_i) \cap (Y = y_i)) - \mathbb{E}\,(X) \times \mathbb{E}\,(Y) \quad (6.159)$$

On vérifie que cette covariance est nulle puisque les variables sont indépendantes.

$$\mathbb{C}ov\,(X,Y) = 0{,}14 \times 10 + 0{,}06 \times 20 + 0{,}56 \times 20 + 0{,}24 \times 40 - 1{,}3 \times 18 = 0 \quad (6.160)$$

5.2.2 Matrice de variance-covariance

Comme nous l'avons dit, une autre façon de présenter un couple de variables aléatoires consiste à définir un vecteur de variables aléatoires $Z = (X,Y)^\top$ de dimension 2×1. Les moments de ce vecteur sont alors des vecteurs ou des matrices. En particulier, l'espérance est un vecteur de dimension 2×1 tel que :

$$\underset{(2\times1)}{\mathbb{E}\,(Z)} = \mathbb{E}\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \mathbb{E}\,(X) \\ \mathbb{E}\,(Y) \end{pmatrix} \quad (6.161)$$

Le concept de « variance » du vecteur Z correspond à l'espérance du « carré » de l'écart $Z - \mathbb{E}\,(Z)$. Mais puisque $Z - \mathbb{E}\,(Z)$ est un vecteur, cette notion de « carré » est remplacée par le produit vectoriel $(Z - \mathbb{E}\,(Z))\,(Z - \mathbb{E}\,(Z))^\top$. Ainsi, la « variance » du

vecteur Z devient :

$$\underset{(2\times2)}{\mathbb{V}(Z)} = \mathbb{E}\left((\underset{(2\times1)}{Z - \mathbb{E}(Z)})\times(\underset{(1\times2)}{Z - \mathbb{E}(Z)})^{\top}\right) \quad (6.162)$$

En développant ces termes, on obtient :

$$\underset{(2\times2)}{\mathbb{V}(Z)} = \mathbb{E}\left(\begin{pmatrix} X - \mathbb{E}(X) \\ Y - \mathbb{E}(Y) \end{pmatrix}\begin{pmatrix} X - \mathbb{E}(X) & Y - \mathbb{E}(Y) \end{pmatrix}\right) \quad (6.163)$$

$$= \begin{pmatrix} \mathbb{E}\left((X - \mathbb{E}(X))^2\right) & \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) \\ \mathbb{E}((Y - \mathbb{E}(Y))(X - \mathbb{E}(X))) & \mathbb{E}\left((Y - \mathbb{E}(Y))^2\right) \end{pmatrix}$$

ou encore :

$$\underset{(2\times2)}{\mathbb{V}(Z)} = \begin{pmatrix} \mathbb{V}(X) & \mathbb{C}ov(X,Y) \\ \mathbb{C}ov(X,Y) & \mathbb{V}(Y) \end{pmatrix} \quad (6.164)$$

On obtient ainsi une **matrice de variance-covariance** dont les termes de la diagonale principale sont les variances des composantes du vecteur Z (*i.e.* les variables X et Y) et les termes hors-diagonale correspondent aux covariances.

De façon générale, pour un vecteur de variables aléatoires $X = (X_1,...,X_k)^{\top}$ de dimension $k \times 1$, on peut définir un *vecteur espérance* de dimension $k \times 1$ et une *matrice de variance-covariance* de dimension $k \times k$.

Définition 6.39

Soit $X = (X_1,...,X_k)^{\top}$ un vecteur de variables aléatoires de dimension $k \times 1$, son **espérance** et sa **matrice de variance-covariance** sont définies par :

$$\underset{(k\times1)}{\mathbb{E}(X)} = \mathbb{E}\begin{pmatrix} X_1 \\ ... \\ X_k \end{pmatrix} = \begin{pmatrix} \mathbb{E}(X_1) \\ ... \\ \mathbb{E}(X_k) \end{pmatrix} \quad (6.165)$$

$$\underset{(k\times k)}{\mathbb{V}(X)} = \begin{pmatrix} \mathbb{V}(X_1) & \mathbb{C}ov(X_1,X_2) & ... & \mathbb{C}ov(X_1,X_i) & ... & \mathbb{C}ov(X_1,X_k) \\ \mathbb{C}ov(X_2,X_1) & \mathbb{V}(X_2) & ... & \mathbb{C}ov(X_2,X_i) & ... & \mathbb{C}ov(X_2,X_k) \\ ... & ... & ... & ... & ... & ... \\ \mathbb{C}ov(X_i,X_1) & \mathbb{C}ov(X_i,X_2) & ... & \mathbb{V}(X_i) & ... & \mathbb{C}ov(X_i,X_k) \\ ... & ... & ... & ... & ... & ... \\ \mathbb{C}ov(X_k,X_1) & \mathbb{C}ov(X_k,X_2) & ... & \mathbb{C}ov(X_k,X_i) & ... & \mathbb{V}(X_k) \end{pmatrix}$$

Propriété

Matrice de variance-covariance

Une matrice de variance-covariance est une matrice carrée, symétrique et définie-positive.

Rappelons qu'une matrice A est symétrique si $A^{\top} = A$. La propriété de symétrie de la matrice de variance-covariance tient au fait que $\mathbb{C}ov\left(X_i,X_j\right) = \mathbb{C}ov\left(X_j,X_i\right)$. La matrice de variance-covariance est en outre définie-positive, cela signifie que toutes ses valeurs propres sont positives strictement. Rappelons que dans le cas scalaire, $k = 1$, la variance ne peut pas être négative, ni nulle. Dans le cas matriciel, cela implique que pour tout vecteur Z non nul de dimension $k \times 1$, le scalaire correspondant à la *forme quadratique* $Z^{\top}\mathbb{V}(X)Z$ est strictement positif.

5.3 Loi conditionnelle

Le concept de probabilité conditionnelle a été défini ci-avant (▶ chapitre 5). Dans le cadre des variables aléatoires, ce concept se traduit par la notion de **loi de probabilité conditionnelle** (ou distribution conditionnelle).

5.3.1 Cas d'un couple de variables discrètes

Soient deux variables aléatoires discrètes X et Y respectivement définies sur $X(\Omega)$ et $Y(\Omega)$. Caractérisons la loi de probabilité conditionnelle (ou distribution conditionnelle) de la variable X sachant que $Y = y$, où $y \in Y(\Omega)$.

Définition 6.40

L'application $\Pr(X = x_i | Y = y_i)$, $\forall x_i \in X(\Omega)$ définit la **loi de probabilité conditionnelle** de la variable X sachant $Y = y_i$. Par définition :

$$\Pr(X = x_i | Y = y_i) = \frac{\Pr((X = x_i) \cap (Y = y_i))}{\Pr(Y = y_i)} \quad \forall x_i \in X(\Omega) \qquad (6.166)$$

Le terme $\Pr(X = x_i | Y = y_i)$ se lit « probabilité que la variable X prenne la valeur x_i sachant que la variable Y est égale à y_i ». Cette définition peut être interprétée comme une application du théorème de Bayes (▶ chapitre 5) au système complet des réalisations (x_i,y_i). De façon symétrique, on peut définir la loi de probabilité de la variable Y sachant $X = x_i$ par :

$$\Pr(Y = y_i | X = x_i) = \frac{\Pr((X = x_i) \cap (Y = y_i))}{\Pr(X = x_i)} \quad \forall y_i \in Y(\Omega) \qquad (6.167)$$

Les probabilités conditionnelles, comme toute probabilité, somment à l'unité.

$$\sum_{x_i \in X(\Omega)} \Pr(X = x_i | Y = y_i) = \frac{1}{\Pr(Y = y_i)} \sum_{x_i \in X(\Omega)} \Pr((X = x_i) \cap (Y = y_i))$$
$$= \frac{\Pr(Y = y_i)}{\Pr(Y = y_i)} = 1 \qquad (6.168)$$

Exemple

On considère deux variables aléatoires indépendantes X et Y respectivement définies sur $X(\Omega) = \{a,b\}$ et $Y(\Omega) = \{1,2\}$. On admet que :

$$\Pr(X = a) = 0{,}2 \qquad \Pr(Y = 1) = 0{,}7 \qquad (6.169)$$
$$\Pr((X = a) \cap (Y = 1)) = 0{,}14 \quad \Pr((X = b) \cap (Y = 1)) = 0{,}56 \qquad (6.170)$$

La loi conditionnelle de X sachant que $Y = 1$ est définie par les probabilités suivantes :

$$\Pr(X = a | Y = 1) = \frac{\Pr((X = a) \cap (Y = 1))}{\Pr(Y = 1)} = \frac{0{,}14}{0{,}7} = 0{,}20 \qquad (6.171)$$

$$\Pr(X = b | Y = 1) = \frac{\Pr((X = b) \cap (Y = 1))}{\Pr(Y = 1)} = \frac{0{,}56}{0{,}7} = 0{,}80 \qquad (6.172)$$

Dans l'exemple précédent, on vérifie que $\Pr(X = a | Y = 1) = \Pr(X = a)$ et que $\Pr(X = b | Y = 1) = \Pr(X = b)$. Les probabilités conditionnelles sont égales aux probabilités marginales. Dit autrement, le fait de savoir que $Y = 1$ ne modifie en rien la loi

de probabilité de X. Cette propriété est la conséquence de l'hypothèse d'indépendance des variables X et Y.

Propriété

Indépendance

Les variables aléatoires discrètes X et Y sont **indépendantes** lorsque les relations suivantes sont vérifiées $\forall x_i \in X(\Omega)$ et $\forall y_i \in Y(\Omega)$:

$$\Pr((X = x_i) \cap (Y = y_i)) = \Pr(X = x_i) \times \Pr(Y = y_i) \tag{6.173}$$

$$\Pr(X = x_i | Y = y_i) = \Pr(X = x_i) \tag{6.174}$$

$$\Pr(Y = y_i | X = x_i) = \Pr(Y = y_i) \tag{6.175}$$

En cas d'indépendance, les probabilités jointes sont égales au produit des probabilités marginales et les probabilités conditionnelles sont égales aux probabilités marginales. Rappelons que l'indépendance est une notion conditionnelle à une certaine mesure de probabilité (► chapitre 5).

Les moments associés à la loi de probabilité conditionnelle ou « moments conditionnels » sont définis de la façon suivante.

Définition 6.41

Pour un univers des réalisations fini $X(\Omega) = \{x_1, \ldots, x_n\}$, les **moments conditionnels** ordinaires et centrés de la variable X sachant $Y = y_i$ sont définis par :

$$\mathbb{E}\left(X^k \middle| Y = y_i\right) = \sum_{i=1}^{n} x_i^k \Pr(X = x_i | Y = y_i) \tag{6.176}$$

$$\mathbb{E}\left((X - \mathbb{E}(X))^k \middle| Y = y_i\right) = \sum_{i=1}^{n} (x_i - \mathbb{E}(X))^k \Pr(X = x_i | Y = y_i) \tag{6.177}$$

En particulier on peut définir l'**espérance conditionnelle** (moment ordinaire d'ordre un) et la **variance conditionnelle** (moment centré d'ordre deux) comme suit :

$$\mathbb{E}(X | Y = y_i) = \sum_{i=1}^{n} x_i \Pr(X = x_i | Y = y_i) \tag{6.178}$$

$$\mathbb{V}(X | Y = y_i) = \sum_{i=1}^{n} (x_i - \mathbb{E}(X))^2 \Pr(X = x_i | Y = y_i) \tag{6.179}$$

Toutes ces définitions peuvent être généralisées au cas d'un *vecteur* de variables aléatoires discrètes.

5.3.2 Cas d'un couple de variables continues

Soient deux variables aléatoires continues X et Y respectivement définies sur $X(\Omega)$ et $Y(\Omega)$. La loi de probabilité conditionnelle (ou distribution conditionnelle) de la variable X sachant que $Y = y$, où $y \in Y(\Omega)$, est définie de la façon suivante.

Définition 6.42

Soit (X,Y) un couple de variables aléatoires réelles continues définies sur le support $X(\Omega) \times Y(\Omega) \subseteq \mathbb{R}^2$. La fonction de **densité conditionnelle** de la variable X sachant que $Y = c$ est définie par :

$$f_{X|Y}(x|c) = \frac{f_{X,Y}(x,c)}{\int_{-\infty}^{+\infty} f_{X,Y}(x,c)\,dx} = \frac{f_{X,Y}(x,c)}{f_Y(c)} \quad \forall x \in X(\Omega) \qquad (6.180)$$

où $f_{X,Y}(x,c)$ désigne la densité jointe et $f_Y(c)$ la densité marginale de la variable Y évaluée pour une valeur c.

La densité conditionnelle peut être notée de façon équivalente par :

$$f_{X|Y}(x|c) \equiv f_{X|Y=c}(x) \equiv f_{X|c}(x) \qquad (6.181)$$

Exemple

On considère un couple de variables aléatoires réelles continues et indépendantes (X,Y) définies sur $\mathbb{R}^2$, admettant une distribution jointe *normale bivariée* standard telle que :

$$f_{X,Y}(x,y) = \frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right) \quad \forall (x,y) \in \mathbb{R}^2 \qquad (6.182)$$

On admet que la densité marginale de la variable Y est égale à :

$$f_Y(y) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right) \quad \forall y \in \mathbb{R} \qquad (6.183)$$

La densité conditionnelle de X sachant que $Y = y$ est définie par :

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right)\right]^{-1}\left(\frac{1}{2\pi}\exp\left[-\frac{x^2+y^2}{2}\right)\right] \qquad (6.184)$$

On obtient donc que :

$$f_{X|Y}(x|y) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right) \quad \forall x \in \mathbb{R} \qquad (6.185)$$

Cette fonction de densité correspond à celle d'une loi normale centrée réduite. Puisque les variables X et Y sont indépendantes, la densité conditionnelle de X sachant $Y = y$ correspond à la densité marginale de X.

Propriété

Indépendance

Les variables aléatoires continues X et Y sont **indépendantes** lorsque les relations suivantes sont vérifiées $\forall x \in X(\Omega)$ et $\forall y \in Y(\Omega)$:

$$f_{X,Y}(x,y) = f_X(x) \times f_Y(y) \qquad (6.186)$$

$$f_{X|Y}(x|y) = f_X(x) \qquad (6.187)$$

$$f_{Y|X}(y|x) = f_Y(y) \qquad (6.188)$$

Sous l'hypothèse d'indépendance, la densité jointe est égale au produit des densités marginales et les densités conditionnelles correspondent aux densités marginales.

Définition 6.43

Les **moments conditionnels** ordinaires et centrés de la variable X sachant $Y = y_i$ sont définis par :

$$\mathbb{E}\left(X^k \middle| Y = y_i\right) = \int_{-\infty}^{+\infty} x^k\ f_{X|Y}(x|y)\ dx \tag{6.189}$$

$$\mathbb{E}\left((X - \mathbb{E}(X))^k \middle| Y = y_i\right) = \int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^k\ f_{X|Y}(x|y)\ dx \tag{6.190}$$

L'**espérance conditionnelle** (moment ordinaire d'ordre un) et la **variance conditionnelle** (moment centré d'ordre deux) vérifient :

$$\mathbb{E}(X|Y = y_i) = \int_{-\infty}^{+\infty} x^k\ f_{X|Y}(x|y)\ dx \tag{6.191}$$

$$\mathbb{V}(X|Y = y_i) = \int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^k\ f_{X|Y}(x|y)\ dx \tag{6.192}$$

Il convient de noter que la détermination de ces moments conditionnels ne requiert pas nécessairement la connaissance de la densité conditionnelle. Par exemple, dans le cadre d'un modèle linéaire du type $Y = a + bX$, dès lors que l'on sait que $X = x$, on peut traiter X comme une constante dans le calcul de l'espérance et de la variance conditionnelles de la variable Y.

Exemple

Soit un modèle de régression linéaire tel que

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{6.193}$$

où Y désigne une variable dépendante, X une variable explicative et ε un terme d'erreur aléatoire. On ne connaît pas la loi de ε, mais l'on suppose que $\mathbb{E}(\varepsilon|X = x) = 0$ et $\mathbb{V}(\varepsilon|X = x) = \sigma^2$. La variable explicative X est distribuée selon une loi inconnue. Déterminons l'espérance et la variance conditionnelles de Y sachant $X = x$.

$$\mathbb{E}(Y|X = x) = \mathbb{E}(\beta_0 + \beta_1 X + \varepsilon|X = x) \tag{6.194}$$

$$= \beta_0 + \beta_1 x + \mathbb{E}(\varepsilon|X = x) \tag{6.195}$$

$$= \beta_0 + \beta_1 x \tag{6.196}$$

De la même façon, on obtient :

$$\mathbb{V}(Y|X = x) = \mathbb{V}(\beta_0 + \beta_1 X + \varepsilon|X = x) \tag{6.197}$$

$$= \mathbb{V}((\beta_0 + \beta_1 x) + (\varepsilon|X = x)) \tag{6.198}$$

$$= \mathbb{V}(\varepsilon|X = x) = \sigma^2 \tag{6.199}$$

Ainsi nous sommes capables de caractériser $\mathbb{E}(Y|X = x)$ et $\mathbb{V}(Y|X = x)$ sans connaître la loi conditionnelle de Y sachant $X = x$ et sa densité conditionnelle.



3 questions à

Stéphanie **Tring**

Chargée d'études statistiques chez AXA Direct Protection



Quel est votre parcours professionnel et votre mission actuelle chez AXA ?

À l'issue de mon stage de fin d'études du master ESA effectué chez BNP Personal Finance, j'ai été embauchée en 2013 chez AXA Direct Protection. Au sein de la direction développement produit et de la planification stratégique, je suis chargée de maximiser la connaissance clients et prospects. Cela se traduit par du ciblage client et de l'analyse du parcours du client. Pour cela j'utilise principalement des méthodes de segmentation et de scoring. Ma mission consiste également à suivre et à analyser les campagnes marketing implémentées par le marketing opérationnel et le marketing digital. L'analyse des campagnes passe, entre autres, par l'étude des profils de clients en situation d'impayé et l'analyse de la persistance dans le portefeuille à plusieurs horizons.

Dans le cadre de votre activité professionnelle, quelle est l'utilité pratique du concept de variable aléatoire ?

Notre activité quotidienne est fondée sur une représentation des caractéristiques ou des comportements des clients par le biais de variables aléatoires. Par exemple, dans le cadre des modélisations de score, on souhaite mesurer l'appétence (représentée par une variable aléatoire) d'un prospect pour un produit à partir de variables socio-démographiques (variables aléatoires). Les variables utilisées pour la modélisation sont qualitatives ou quantitatives.

Quelle est le rôle de la statistique dans les activités marketing d'un groupe comme AXA ?

Au sein du marketing, la statistique permet d'avoir une meilleure connaissance client et prospect, afin de proposer le produit le plus adapté à chaque client. De plus, l'utilisation de la statistique a également un impact sur les coûts d'acquisition des clients qui peuvent être réduits grâce à une meilleure connaissance du portefeuille. Le rôle du chargé d'études est aussi d'approfondir certains sujets particuliers tels que les impayés ou la diminution de l'érosion du portefeuille clients. ■

Les points **clés**

- Une variable aléatoire est une application mesurable d'un univers des possibles probabilisé vers un univers des réalisations probabilisables.
- Le support de la distribution d'une variable aléatoire correspond à l'univers de ses réalisations.
- Une variable aléatoire discrète est définie sur un support fini ou infini dénombrable.
- Une variable aléatoire continue est définie sur un support infini non dénombrable.
- La fonction de masse d'une variable aléatoire discrète correspond à la probabilité associée à une réalisation particulière.
- Pour une variable continue, la probabilité associée à une réalisation particulière est nulle.
- La fonction de répartition correspond à la probabilité cumulée que les réalisations d'une variable aléatoire (discrète ou continue) soient inférieures à une certaine valeur.
- Un quantile est défini par l'inverse de la fonction de répartition.
- Une loi de probabilité discrète ou continue peut être caractérisée par la fonction de masse ou de densité suivant les cas, par la fonction de répartition ou par la population des moments (ordinaires ou centrés).
- L'espérance, moment ordinaire d'ordre un, est un opérateur linéaire.
- La variance, moment centré d'ordre deux, est un opérateur quadratique.
- Pour un couple ou un vecteur de variables aléatoires, on distingue les notions de distribution marginale, distribution jointe et distribution conditionnelle.

ÉVALUATION

► Corrigés sur www.dunod.com

QCM

Pour chacune des questions suivantes, indiquer si les affirmations sont vraies ou fausses (il peut y avoir plusieurs réponses vraies pour chaque question).

1 Variable aléatoire

a. Une variable aléatoire est une application.

b. Une variable aléatoire qualitative peut être continue.

c. Une variable aléatoire est définie sur un univers probabilisé.

d. Une variable aléatoire continue est définie sur un support infini.

e. Une variable aléatoire discrète est définie sur un support infini.

2 Fonction de densité, de masse et de répartition

a. La fonction de masse correspond à une probabilité.

b. Une densité est comprise entre 0 et 1.

c. La fonction de densité est la primitive de la fonction de répartition.

d. La fonction de répartition correspond à une probabilité cumulée.

e. La fonction de masse est la dérivée de la fonction de répartition.

3 Fonction de répartition et quantile

a. Un quantile est une probabilité.

b. La fonction de répartition inverse est croissante sur [0,1].

c. Si le support de la loi est une partie de $\mathbb{R}$, les quantiles sont définis sur $\mathbb{R}$.

d. La fonction de répartition a toujours une expression analytique.

e. La fonction de répartition est croissante sur le support de la loi de probabilité.

4 Indépendance

a. Si la covariance entre deux variables est nulle, ces variables sont indépendantes.

b. Si la densité jointe est égale au produit des densités conditionnelles, les variables sont indépendantes.

c. Si les densités conditionnelles sont égales aux densités marginales, les variables sont indépendantes.

d. Si deux variables sont indépendantes, leur corrélation est nulle.

e. La notion d'indépendance est relative à une mesure de probabilité.

Exercices

5 Fonction de répartition

Soit X une variable aléatoire réelle. On suppose que sa fonction de répartition $F_X(x)$ est donnée par :

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1/4 & \text{si } 0 \leq x < 1 \\ 3/4 & \text{si } 1 \leq x < 2 \\ 1 & \text{si } x \geq 2 \end{cases} \qquad (6.200)$$

1. Calculer la probabilité $\Pr(-1/2 < X < 1/2)$.

2. Calculer la probabilité $\Pr(-1/2 < X < 3/2)$.

3. Calculer la probabilité $\Pr(X > 3)$.

6 Espérance

Soit X une variable aléatoire et a un nombre réel quelconque.

1. Démontrer que :

$$\mathbb{E}((X-a)^2) = \mathbb{V}(X) + (\mathbb{E}(X-a))^2 \qquad (6.201)$$

où $\mathbb{E}(.)$ désigne l'espérance et $\mathbb{V}(.)$ désigne la variance.

2. Déterminer la valeur de a pour laquelle l'espérance $\mathbb{E}((X-a)^2)$ est minimum.

7 Fonction de densité

Soit θ un nombre réel et $f_X(.)$ une fonction définie par :

$$f_X(x) = \begin{cases} 0 & \text{si } x \leq \theta \\ \exp(-(x-\theta)) & \text{si } x > \theta \end{cases} \quad (6.202)$$

1. Montrer que $f_X(.)$ satisfait aux conditions requises pour être la densité de probabilité d'une variable aléatoire continue. On utilisera pour cela la fonction gamma, notée $\Gamma(z)$, telle que :

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} \exp(-t)\, dt \quad \forall z \in \mathbb{R}^+ \quad (6.203)$$

$$\Gamma(z) = (z-1)! \quad \forall z \in \mathbb{N} \quad (6.204)$$

2. Déterminer la fonction de répartition $F_X(x)$ de la variable aléatoire X de densité $f_X(.)$.
3. Exprimer l'espérance $\mathbb{E}(X)$ en fonction de la valeur de θ.
4. Exprimer le moment simple d'ordre 2 de la variable aléatoire X et sa variance $\mathbb{V}(X)$ en fonction de la valeur de θ.

Sujets d'examen

8 Transformée de variable aléatoire (d'après Edhec 2009, voie E)

Dans cet exercice, p désigne un réel de $]0,1[$ et on note $q = 1 - p$. On considère deux variables aléatoires X et Y définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \Pr)$, indépendantes et suivant toutes deux la même loi géométrique de paramètre p telle que :

$$\Pr(X = k) = \Pr(Y = k) = q^{k-1} p \quad \forall k \in \mathbb{N}^* \quad (6.205)$$

On pose :

$$Z = \inf(X, Y) \quad (6.206)$$

et on admet que Z est une variable aléatoire, elle aussi définie sur le même espace probabilisé $(\Omega, \mathcal{F}, \Pr)$. On rappelle que pour tout entier naturel k, on a l'égalité :

$$(Z > k) = (X > k) \cap (Y > k) \quad (6.207)$$

1. Pour tout entier naturel k, calculez $\Pr(Z > k)$.
2. Établir que, pour tout entier naturel k supérieur ou égal à 1, on a :

$$\Pr(Z = k) = \Pr(Z > k-1) - \Pr(z > k) \quad (6.208)$$

3. En déduire que Z suit une loi géométrique de paramètre $\left(1 - q^2\right)$.

9 Variable aléatoire discrète (d'après EM Lyon, voie E)

On dispose d'un jeu de $2n$ cartes, avec $n \in \mathbb{N}^*$, qui contient deux rois rouges. Les cartes du jeu sont alignées sur une table de façon aléatoire. Le joueur retourne les cartes une par une jusqu'à obtenir un roi rouge. On définit l'événement E_k comme « le premier roi rouge obtenu est la $k^{ème}$ carte retournée ».

1. Calculer $\Pr(E_1)$, puis en fonction de n et de k définir la probabilité $\Pr(E_k)$ pour $k \geq 2$.
2. Le joueur donne un euro à chaque carte retournée et dès qu'il obtient un roi rouge, il obtient a euros et le jeu s'arrête. Son gain est représenté par la variable aléatoire X. Quelle est la valeur de X si le premier roi rouge est la $k^{ème}$ carte retournée ?
3. Démontrer que $\forall k \in \{1, \ldots, 2n\}$:

$$\Pr(X = a - k) = \frac{2n - k}{n(2n-1)} \quad (6.209)$$

4. Vérifier que :

$$\sum_{k=1}^{2n} \Pr(X = a - k) = 1 \quad (6.210)$$

10 Variable aléatoire continue

On considère une fonction $f_X(x)$ définie par :

$$f_X(x) = a \exp(-|x|) \quad \forall x \in \mathbb{R} \quad (6.211)$$

où a est une constante réelle.

1. Déterminer la constante a pour que la fonction $f_X(x)$ soit la fonction de densité d'une variable aléatoire réelle X.
2. Déterminer la fonction de répartition de la variable aléatoire X.
3. Calculer l'espérance de la variable aléatoire X.

Chapitre 7

Certaines lois de probabilité possèdent des propriétés particulières et sont très souvent employées pour modéliser les phénomènes de la vie quotidienne ou de la vie économique. Du fait de leur utilisation fréquente, on les qualifie de lois de probabilité usuelles ou de lois usuelles. Toutes ces lois sont désignées par un nom[1], par exemple loi binomiale, loi de Poisson, loi de Student, loi binomiale négative, etc. Les lois usuelles, discrètes ou continues, sont souvent des lois paramétriques, cela signifie que leur fonction de masse ou de densité dépend d'un ou de plusieurs paramètres. Par exemple, la fonction de masse associée à une loi de Poisson dépend d'un paramètre positif noté λ. Les noms des lois usuelles sont représentés par des abréviations[2] qui font souvent apparaître les paramètres de leur fonction de densité ou leur fonction de masse. Ainsi, la loi de Poisson est notée $\mathcal{P}(\lambda)$, la loi binomiale $\mathcal{B}(n,p)$, etc.

LES GRANDS AUTEURS



William Gosset (1876-1937)

Les lois de probabilité usuelles portent soit un nom qui rappelle leurs principales propriétés statistiques (loi exponentielle, loi uniforme, etc.), soit le nom des mathématiciens qui les ont inventées (loi de Laplace-Gauss, loi de Bernoulli, loi de Poisson, etc.). La **loi de Student** fait figure d'exception : ce nom ne fait référence à aucune propriété particulière et il n'existe pas de madame ou de monsieur Student. Cette loi fut en fait découverte en 1908 par un statisticien du nom de **William Gosset** qui travaillait à l'époque pour la brasserie Guinness à Dublin.
Dans le but de déterminer une méthode de sélection des meilleures variétés d'orge, il inventa un test statistique (dit test t) et détermina la loi de la statistique de ce test (► chapitre 11). Mais le dirigeant de la société Guinness avait imposé à tous ses employés de ne jamais rien publier, quel que fut le sujet, afin de garder les secrets de fabrication de la brasserie. Il fit toutefois une exception pour William Gosset en lui demandant de prendre un pseudonyme. Le statisticien choisit alors le nom de Student et c'est ainsi que fut baptisée la loi de Student. ■

1 L'utilisation d'une majuscule signifie que le nom de la loi correspond au nom de son découvreur. Par exemple, la loi de Poisson fait référence au mathématicien français Denis Poisson (1781-1840).
2 Ces abréviations sont souvent notées avec une police de type calligraphique, $\mathcal{P}$, $\mathcal{N}$, $\mathcal{F}$, etc.

Lois de probabilité usuelles

Plan

Pré-requis

→ **Connaître** la notion de variable aléatoire (▸ chapitre 6).

Objectifs

→ **Présenter** les principales lois usuelles.

→ **Savoir lire** les tables statistiques des principales lois usuelles.

→ **Savoir calculer** une probabilité pour les principales lois usuelles.

Dans ce chapitre, nous présenterons les principales **propriétés** de certaines **lois usuelles** discrètes ou continues : fonction de densité ou fonction de masse, fonction de répartition, quantile, fonction génératrice des moments, moments remarquables, etc. Aucune démonstration ne sera présentée puisque celles-ci peuvent être retrouvées à partir des principes généraux présentés dans le chapitre 6. Nous insisterons plutôt sur le contexte d'application de ces lois.

1 Lois usuelles discrètes

1.1 Loi uniforme discrète

La **loi uniforme discrète** est une loi de probabilité définie sur un support fini pour laquelle toutes les réalisations sont **équiprobables**. Les exemples typiques d'application de cette loi sont ceux du lancer d'un dé parfaitement équilibré, du tirage d'une carte au hasard ou du tirage d'un numéro au hasard dans une loterie.

1.1.1 Fonction de masse et fonction de répartition

Définition 7.1

La variable aléatoire discrète X suit une **loi uniforme discrète** sur le support fini $X(\Omega) = \{x_1, \ldots, x_n\}$, si sa fonction de masse est définie par :

$$f_X(x) = \Pr(X = x) = \frac{1}{n} \quad \forall x \in X(\Omega) \tag{7.1}$$

On vérifie que toutes les réalisations ont la même probabilité : c'est la propriété d'*équiprobabilité* qui caractérise la loi uniforme. Notons que les réalisations peuvent être *quantitatives*, c'est-à-dire correspondre à des nombres (par exemple si $X(\Omega) = \{1,2,3,4,5,6\}$ dans le cas d'un lancer de dé). Elles peuvent être aussi *qualitatives* (si elles ne sont pas des nombres), par exemple si $X(\Omega) = \{$« valet », « dame », « roi »$\}$ dans le cas le cas d'un jeu de cartes à trois cartes. Dans tous les cas, toutes ces réalisations ont la même probabilité de survenue.

Afin de simplifier les notations, nous allons considérer le cas où la variable X est définie sur un ensemble d'entiers consécutifs $X(\Omega) = \{a, a+1, \ldots, b-1, b\}$ avec $n = b - a + 1$. Dans ce cas, la fonction de masse devient :

$$f_X(x) = \Pr(X = x) = \frac{1}{b - a + 1} \tag{7.2}$$

Exemple

Considérons une variable aléatoire X distribuée selon une loi uniforme discrète sur $X(\Omega) = \{1, \ldots, 10\}$, sa fonction de masse est définie par $f_X(x) = 1/10$, $\forall x \in X(\Omega)$. Toutes les réalisations ont la même probabilité.

Définition 7.2

Si la variable aléatoire X admet une loi uniforme discrète sur $X(\Omega) = \{a, a+1, \ldots, b-1, b\}$, sa **fonction de répartition** $F_X(x) = \Pr(X \leq x)$ est définie par $\forall x \in \mathbb{R}$:

$$F_X(x) = \begin{cases} 0 & \text{si } x < a \\ \dfrac{x-a+1}{b-a+1} & \text{si } a \leq x \leq b \\ 1 & \text{si } x > b \end{cases} \quad (7.3)$$

Rappelons qu'une fonction de répartition est toujours définie sur $\mathbb{R}$, y compris dans le cas d'une variable aléatoire discrète (▶ chapitre 6). Cette fonction de répartition se présente sous la forme de *marches d'escalier* sur le segment $[a,b]$, la « hauteur » des marches étant égale à $1/(b-a+1)$, c'est-à-dire $1/10$ dans le cas de notre exemple.

1.1.2 Moments

La fonction génératrice des moments de la loi uniforme discrète sur le support $X(\Omega) = \{a, \ldots, b\}$ est égale à :

$$M_X(t) = \frac{\exp(a \times t)}{b-a+1} \sum_{i=1}^{b-a+1} \exp(i \times t) \quad \forall t \in \mathbb{R} \quad (7.4)$$

De cette fonction génératrice, on peut dériver l'espérance et la variance.

Propriété

Espérance et variance de la loi uniforme discrète

Si X admet une loi uniforme discrète sur $X(\Omega) = \{a, a+1, \ldots, b-1, b\}$, alors :

$$\mathbb{E}(X) = \frac{a+b}{2} \qquad \mathbb{V}(X) = \frac{(b-a+1)^2 - 1}{12} \quad (7.5)$$

1.2 Loi de Bernoulli

La **loi de Bernoulli**, du nom du mathématicien suisse Jacques Bernoulli (1654-1705), est une loi de probabilité discrète définie sur un support fini comportant deux réalisations. C'est la loi que l'on utilise pour représenter des variables aléatoires **dichotomiques** ou **binaires**, c'est-à-dire à deux modalités.

Exemple

La loi de Bernoulli peut être appliquée sur des supports du type $X(\Omega) = \{$« pair », « impair »$\}$, $X(\Omega) = \{$« succès », « échec »$\}$ ou $X(\Omega) = \{10,35\}$.

Ces trois exemples montrent que la loi de Bernoulli peut être appliquée à des variables quantitatives ou à des variables qualitatives. Toutefois, il est toujours possible d'exprimer une variable dichotomique qualitative sous la forme d'une variable quantitative en utilisant un **codage** $(a,b) \in \mathbb{R}^2$. Par exemple, on pose $X = 2$ si « succès » et $X = 3$ si « échec ». Il existe bien évidemment une infinité de codages (a,b) possibles. Par convention, on utilise toujours le codage binaire $(0,1)$. Ainsi, quel que soit le problème modélisé (quantitatif ou qualitatif), on considère l'univers des réalisations $X(\Omega) = \{0,1\}$.

1.2.1 Fonction de masse et fonction de répartition

Définition 7.3

La variable aléatoire discrète X suit une **loi de Bernoulli** si sa fonction de masse est définie par :

$$f_X(x) = \Pr(X = x) = p^x (1-p)^{1-x} \quad \forall x \in X(\Omega) = \{0,1\} \tag{7.6}$$

où le paramètre p est un réel vérifiant $p \in]0,1[$.

Si X suit une loi de Bernoulli de paramètre p, alors on note $X \sim \text{Bernoulli}(p)$ ou $\text{Bern}(p)$. Le paramètre p, appelé **probabilité de succès**[3], correspond à la probabilité que X prenne une réalisation égale à 1, *i.e.* $\Pr(X = 1) = p$.

Définition 7.4

Si la variable aléatoire X admet une loi de Bernoulli de paramètre $p \in]0,1[$, sa **fonction de répartition** $F_X(x) = \Pr(X \leq x)$ est définie par $\forall x \in \mathbb{R}$:

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1-p & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases} \tag{7.7}$$

Comme pour toute variable aléatoire discrète, la fonction de répartition de la loi de Bernoulli se présente sous la forme d'une fonction en marches d'escalier.

1.2.2 Moments

La *fonction génératrice des moments* de la loi de Bernoulli(p) est définie par :

$$M_X(t) = (1-p) + p\exp(t) \quad \forall t \in \mathbb{R} \tag{7.8}$$

De cette fonction génératrice, on peut dériver l'espérance et la variance.

Propriété

Espérance et variance de la loi de Bernoulli

Si X admet une loi de Bernoulli de paramètre $p \in]0,1[$, alors :

$$\mathbb{E}(X) = p \quad \mathbb{V}(X) = p(1-p) \tag{7.9}$$

On remarque que les contraintes sur la probabilité de succès $p \neq 0$ et $p \neq 1$ garantissent que la variance de la variable X soit non nulle. De plus, c'est l'utilisation d'un codage (0,1) pour représenter les variables dichotomiques qui permet d'obtenir l'égalité entre l'espérance de la variable X et la probabilité de succès :

$$\mathbb{E}(X) = p \times 1 + (1-p) \times 0 = p \tag{7.10}$$

3 La valeur 1 est supposée coder le « succès » d'une expérience. Par exemple on code 1 si « pile » et 0 si « face » si l'on s'intéresse au résultat « pile » dans le cadre d'un lancer de pièce.

1.3 Loi binomiale

La **loi binomiale** est une loi de probabilité discrète définie sur le support fini $X(\Omega) = \{0, 1, \ldots, n\}$. Cette loi correspond à une expérience aléatoire dans laquelle on répète n fois de manière indépendante une expérience de Bernoulli avec une probabilité de succès égale à p. On compte alors le nombre de succès, c'est-à-dire le nombre de fois où la réalisation de la variable de Bernoulli est égale à 1. Le nombre total de succès, noté X, est une variable aléatoire admettant une distribution binomiale de paramètres n et p, notée :

$$X \sim \mathcal{B}(n, p) \qquad (7.11)$$

Exemple

Le nombre de résultats « face » apparus lors de n lancers d'une pièce parfaitement équilibrée suit une loi $\mathcal{B}(n, 1/2)$. Le nombre de boules rouges apparues au cours de n tirages avec remise dans une urne contenant 15 boules dont 5 boules rouges suit une loi $\mathcal{B}(n, 1/3)$.

1.3.1 Fonction de masse et fonction de répartition

Définition 7.5

La variable aléatoire discrète X définie sur $X(\Omega) = \{0, \ldots, n\}$ suit une **loi binomiale** $\mathcal{B}(n, p)$ si sa fonction de masse est définie par :

$$f_X(x) = \Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \forall x \in X(\Omega) \qquad (7.12)$$

avec $p \in]0, 1[$ et $n \in \mathbb{N}^*$.

Le paramètre p correspond à la probabilité de succès des épreuves de Bernoulli, il est donc compris entre 0 et 1 exclus. Le paramètre n correspond au nombre de répétitions de l'épreuve de Bernoulli, c'est donc un entier non nul. La fonction de masse de la loi binomiale dépend du nombre de combinaisons de x éléments parmi n. Ce nombre de combinaisons, parfois noté C_n^x (prononcer « x *parmi* n »), est égal à :

$$\binom{n}{x} = C_n^x = \frac{n!}{x!\,(n-x)!} \qquad (7.13)$$

où le symbole « ! » correspond à la factorielle[4].

Exemple

Soit X une variable aléatoire discrète définie sur $X(\Omega) = \{0, 1, 2, 3, 4\}$ et telle que $X \sim \mathcal{B}(4, 1/5)$. On obtient alors :

$$\Pr(X = 0) = \frac{4!}{0! \times 4!} \times 0{,}2^0 \times (1 - 0{,}2)^{4-0} = 0{,}8^4 = 0{,}4096 \qquad (7.14)$$

$$\Pr(X = 1) = \frac{4!}{1! \times 3!} \times 0{,}2^1 \times (1 - 0{,}2)^{4-1} = 4 \times 0{,}2 \times 0{,}512 = 0{,}4096 \qquad (7.15)$$

$$\Pr(X = 2) = \frac{4!}{2! \times 2!} \times 0{,}2^2 \times (1 - 0{,}2)^{4-2} = 6 \times 0{,}04 \times 0{,}64 = 0{,}1536 \qquad (7.16)$$

4 Rappelons que la factorielle $k!$ d'un entier naturel k est le produit des nombres entiers strictement positifs inférieurs ou égaux à k, *i.e.* $k \times (k-1) \times \ldots \times 1$. Par exemple, $3! = 3 \times 2 \times 1$.

Rappelons que la combinaison vérifie les propriétés suivantes :

$$\binom{n}{n} = \binom{n}{0} = 1 \tag{7.17}$$

$$\binom{n}{x} = \binom{n}{n-x} \quad \forall x \in \{0,\ldots,n\} \tag{7.18}$$

Propriété

Loi binomiale

Étant données les propriétés de la combinaison, on en déduit que si $X \sim \mathcal{B}(n,p)$ alors :

$$\Pr(X = 0) = (1-p)^n \qquad \Pr(X = n) = p^n \tag{7.19}$$

Si $X \sim \mathcal{B}(n,p)$ et si $Y \sim \mathcal{B}(n,1-p)$ alors $\forall k \in \{0,\ldots,n\}$:

$$\Pr(X = k) = \Pr(Y = 1-k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{7.20}$$

Définition 7.6

Si la variable aléatoire X admet une loi binomiale $\mathcal{B}(n,p)$, sa **fonction de répartition** $F_X(x) = \Pr(X \leq x)$ est définie par $\forall x \in \mathbb{R}$:

$$F_X(x) = 0 \quad \text{si } x < 0 \tag{7.21}$$

$$F_X(x) = \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k} \quad \text{si } 0 \leq x \leq n \tag{7.22}$$

$$F_X(x) = 1 \quad \text{si } x > n \tag{7.23}$$

où $\lfloor x \rfloor$ désigne[5] le plus grand entier inférieur ou égal à x.

Exemple

Soit X une variable aléatoire discrète définie sur $X(\Omega) = \{0,1,2,3,4\}$ et telle que $X \sim \mathcal{B}(4 ; 0{,}2)$, alors :

$$F_X(1) = \sum_{k=0}^{\lfloor 1 \rfloor} \Pr(X = k) = \Pr(X = 0) + \Pr(X = 1) = 0{,}8192 \tag{7.24}$$

puisque $\lfloor 1 \rfloor = 1$. De la même façon :

$$F_X(1{,}58) = \sum_{k=0}^{\lfloor 1{,}58 \rfloor} \Pr(X = k) = \sum_{k=0}^{1} \Pr(X = k) = F_X(1) = 0{,}8192 \tag{7.25}$$

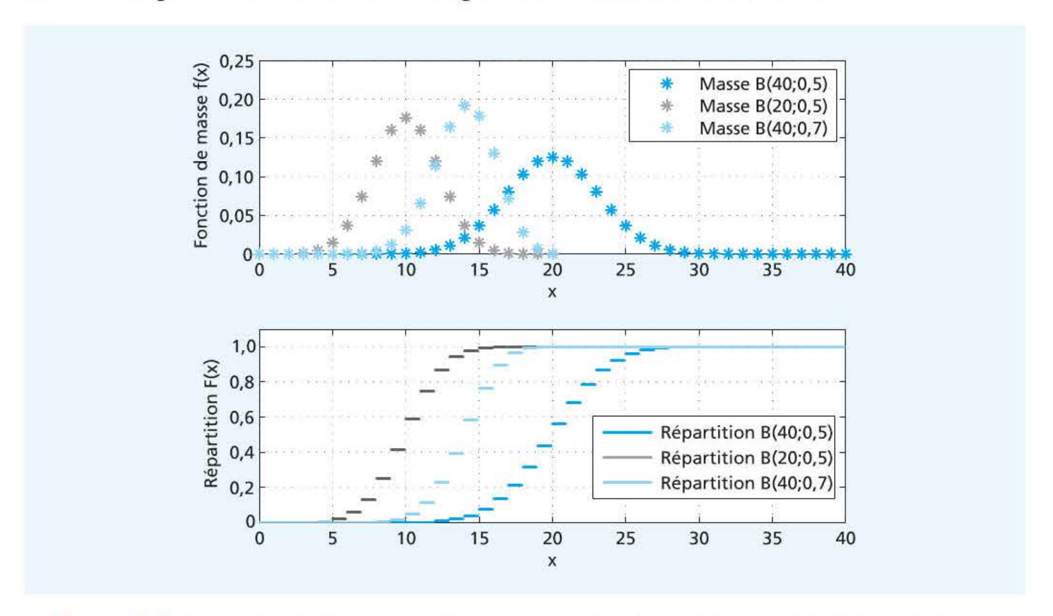
puisque $\lfloor 1{,}58 \rfloor = 1$.

On comprend aisément que le calcul de $F_X(x)$ soit relativement fastidieux notamment lorsque x et/ou n sont grands. C'est pourquoi lorsque l'on souhaite obtenir des probabilités cumulées d'une loi binomiale, on utilise soit les fonctions préprogrammées des logiciels de statistique ou des tableurs (▶ En pratique : la loi binomiale sous Excel), soit des **tables statistiques** de la loi binomiale. Plusieurs tables de la loi binomiale

5 L'opérateur $\lfloor x \rfloor$ est appelée le *floor* (étage en français). Par exemple $\lfloor 1{,}08 \rfloor = 1$ et $\lfloor 1 \rfloor = 1$.

figurent en annexe de cet ouvrage : chaque table correspond à une valeur de n ($n = 10$, $n = 20$, $n = 25$ et $n = 50$). Pour chaque table sont reportées différentes valeurs de la probabilité p (de 0,05 à 0,30). Sur les lignes de chaque table figure une valeur k variant de 0 à n. Pour une valeur de n (une table), une valeur de p (colonne) et une valeur de k (ligne), on trouve la probabilité cumulée $\Pr(X \leq k)$ associée à la loi $\mathcal{B}(n,p)$.

La figure 7.1 représente les fonctions de masse et les fonctions de répartition de trois exemples de lois binomiales $\mathcal{B}(n,p)$ obtenues à partir d'un logiciel statistique. On vérifie que la fonction de masse $\Pr(X = x)$ n'est définie que pour des valeurs de x égales à $0,\ldots,n$. La fonction de répartition est au contraire définie pour toute valeur réelle et se présente sous la forme de plateaux ou de marches d'escalier.



▲ **Figure 7.1** **Exemple de fonctions de masse et de répartition de lois binomiales**

EN PRATIQUE

La loi binomiale sous Excel

Les fonctions de masse et de répartition de la loi binomiale $\mathcal{B}(n,p)$ sont programmées dans tous les logiciels de statistique et d'économétrie et dans la plupart des tableurs. Par exemple sous le tableur Excel, ces fonctions peuvent être appelées en suivant les syntaxes suivantes :

$f_x(x)$: LOI.BINOMIALE(x,n,p,0)

$F_x(x)$: LOI.BINOMIALE(x,n,p,1)

Pour toutes les lois discrètes ou continues, on utilise sous Excel le même type de syntaxe en modifiant bien évidemment le nom de la loi ainsi que la liste des paramètres déclarés.

Le dernier paramètre de la fonction (0 ou 1) sert à renvoyer soit la fonction de masse (ou de densité) pour une valeur 0, soit la fonction de répartition pour une valeur 1.

1.3.2 Moments

La fonction génératrice des moments de la loi binomiale $\mathcal{B}(n,p)$ est définie par :

$$M_X(t) = ((1-p) + p\exp(t))^n \quad \forall t \in \mathbb{R} \tag{7.26}$$

On vérifie que la fonction génératrice des moments d'une loi $\mathcal{B}(n,p)$ correspond à la fonction génératrice des moments d'une loi de Bernoulli, élevée à la puissance n. De cette fonction génératrice, on peut notamment dériver l'espérance et la variance de la loi binomiale.

Propriété

Espérance et variance de la loi binomiale
Si X admet une loi binomiale $\mathcal{B}(n,p)$, alors :

$$\mathbb{E}(X) = np \qquad \mathbb{V}(X) = np(1-p) \tag{7.27}$$

On constate que l'espérance et la variance d'une loi binomiale $\mathcal{B}(n,p)$ sont égales à l'espérance et la variance d'une loi de Bernoulli multipliées par n.

1.3.3 Autres propriétés

Propriété

Somme de variables de Bernoulli indépendantes
Soient $Z_1,\ldots,Z_n$ des variables de Bernoulli(p) indépendantes avec $p \in]0,1[$, alors :

$$\sum_{i=1}^{n} Z_i \sim \mathcal{B}(n,p) \tag{7.28}$$

Cette propriété implique que la loi binomiale est **additive** : la somme de deux variables indépendantes distribuées selon des lois binomiales de même probabilité de succès suit, elle-aussi, une loi binomiale.

Propriété

Additivité de la loi binomiale
Soient X et Y deux variables aléatoires discrètes indépendantes telles que $X \sim \mathcal{B}(n,p)$ et $Y \sim \mathcal{B}(m,p)$, alors

$$X + Y \sim \mathcal{B}(n+m,p) \tag{7.29}$$

Même si, à ce stade du chapitre, nous n'avons pas encore présenté la loi normale (► section 2.3), il convient de mentionner un résultat très souvent employé concernant l'**approximation** de la loi binomiale par *la loi normale* sous certaines conditions sur les paramètres n et p. Lorsque n est suffisamment grand, si $X \sim \mathcal{B}(n,p)$ alors :

$$X \approx \mathcal{N}(np, np(1-p)) \tag{7.30}$$

où le symbole $\approx$ signifie « approximativement distribué selon » et le symbole $\mathcal{N}$ désigne la loi normale. Cette approximation continue est d'autant meilleure que n est grand et que la probabilité p est éloignée des valeurs extrêmes 0 et 1. Il existe plusieurs règles alternatives sur le couple (p,n) pour savoir si l'approximation par la loi normale est adaptée ou non. Nous en proposons une assez simple.

Propriété

Approximation par la loi normale

Soit X une variable aléatoire telle que $X \sim \mathcal{B}(n,p)$. Si $n > 5$ et si :

$$\frac{1}{\sqrt{n}}\left|\left(\sqrt{\frac{1-p}{p}} - \sqrt{\frac{p}{1-p}}\right)\right| < 0{,}3 \tag{7.31}$$

alors l'approximation de la loi binomiale par la loi normale peut être appliquée :

$$X \approx \mathcal{N}(np, np(1-p)) \tag{7.32}$$

1.4 Loi géométrique

La **loi géométrique** est une loi de probabilité discrète pouvant être définie soit sur l'ensemble des entiers $\mathbb{N}$, soit sur l'ensemble des entiers non nuls $\mathbb{N}^*$. Lorsqu'elle est définie sur $\mathbb{N}^*$, la loi géométrique de paramètre p correspond à l'expérience aléatoire suivante. On répète de manière indépendante une expérience de Bernoulli avec une probabilité de succès égale à p jusqu'au premier succès. Soit X la variable qui correspond au rang du premier succès : ce rang est nécessairement supérieur ou égal à 1 et inférieur ou égal à n, donc $X \in X(\Omega) = \{1,2,\ldots,n,\ldots\}$. La variable X admet une distribution géométrique de paramètre p notée :

$$X \sim Geom(p) \quad \text{ou} \quad X \sim \mathcal{G}(p) \tag{7.33}$$

Lorsqu'elle est définie sur $\mathbb{N}$, la loi géométrique correspond à la distribution du nombre d'échecs $Y = X - 1$ avant le premier succès. Le nombre d'échecs peut être égal à 0 en cas de réussite à la première expérience de Bernoulli. La variable $Y \in Y(\Omega) = \{0,1,\ldots,n,\ldots\}$ est distribuée selon une loi géométrique de paramètre p, notée de la même façon $Y \sim \mathcal{G}(p)$. Il convient donc de faire attention au support de la loi géométrique afin d'éviter les confusions.

1.4.1 Fonction de masse et fonction de répartition

Définition 7.7

La variable aléatoire discrète X définie sur $X(\Omega) = \mathbb{N}$ suit une **loi géométrique** $\mathcal{G}(p)$ si sa fonction de masse est définie par :

$$f_X(x) = \Pr(X = x) = (1-p)^x\, p \quad \forall x \in \mathbb{N} \tag{7.34}$$

Si cette variable est définie sur $X(\Omega) = \mathbb{N}^*$, sa fonction de masse devient :

$$f_X(x) = \Pr(X = x) = (1-p)^{x-1}\, p \quad \forall x = \mathbb{N}^* \tag{7.35}$$

où la probabilité de succès p est un réel vérifiant $p \in]0,1]$. On remarque que la loi géométrique peut être définie pour une probabilité de succès p égale à 1, mais pas pour une probabilité nulle.

Définition 7.8

Si la variable aléatoire X admet une loi géométrique $\mathcal{G}(p)$ sur $X(\Omega) = \mathbb{N}$, sa **fonction de répartition** $F_X(x) = \Pr(X \leq x)$ est définie par $\forall x \in \mathbb{R}$:

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - (1-p)^{\lfloor x \rfloor + 1} & \text{si } x \geq 0 \end{cases} \quad (7.36)$$

Si cette variable est définie sur sur $X(\Omega) = \mathbb{N}^*$, sa fonction de répartition devient :

$$F_X(x) = \begin{cases} 0 & \text{si } x < 1 \\ 1 - (1-p)^{\lfloor x \rfloor} & \text{si } x \geq 1 \end{cases} \quad (7.37)$$

où $\lfloor x \rfloor$ désigne le plus grand entier inférieur ou égal à x.

Exemple

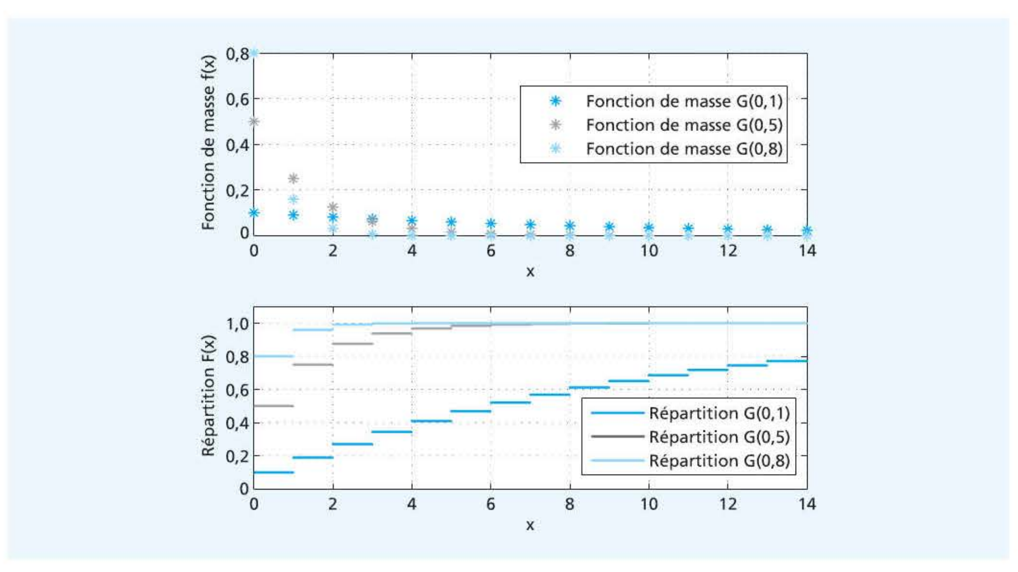
Soit Y une variable aléatoire discrète définie sur $Y(\Omega) = \mathbb{N}$ telle que $X \sim \mathcal{G}(0{,}10)$, alors :

$$\Pr(X \leq 1) = F_X(1) = 1 - (1 - 0{,}1)^{\lfloor 1 \rfloor + 1} = 1 - 0{,}9^2 = 0{,}19 \quad (7.38)$$

$$\Pr(X \leq 1{,}2) = F_X(1{,}2) = 1 - (1 - 0{,}1)^{\lfloor 1{,}2 \rfloor + 1} = 1 - 0{,}9^2 = 0{,}19 \quad (7.39)$$

Pour les deux définitions de la fonction de répartition on vérifie toujours que $\lim_{x \to \infty} F_X(x) = 1$.

La figure 7.2 représente les fonctions de masse et de répartition de trois exemples de lois géométriques $\mathcal{G}(p)$ définies sur $\mathbb{N}$ pour des probabilités de succès $p = 0{,}1$, $p = 0{,}5$ et $p = 0{,}8$. De façon générale, la fonction de masse d'une loi géométrique est décroissante, puisque par exemple pour $X(\Omega) = \mathbb{N}$, on a $\Pr(X = 0) = p$, $\Pr(X = 1) = p(1-p)$, $\Pr(X = 2) = p(1-p)^2$, etc.



▲ **Figure 7.2 Exemple de fonctions de masse et de répartition de lois géométriques**

1.4.2 Moments

La fonction génératrice des moments de la loi $\mathcal{G}(p)$ définie sur $\mathbb{N}$ est égale à :

$$M_X(t) = \frac{p}{1-(1-p)\exp(t)} \quad \forall t \in \mathbb{R} \tag{7.40}$$

Dans le cas où la loi géométrique est définie sur $\mathbb{N}^*$, cette fonction devient :

$$M_X(t) = \frac{p\exp(t)}{1-(1-p)\exp(t)} \quad \forall t < -\ln(1-p) \tag{7.41}$$

De cette fonction génératrice, on peut notamment dériver l'espérance et la variance de la loi géométrique.

Propriété

Espérance et variance de la loi géométrique
Si X admet une loi géométrique $\mathcal{G}(p)$ définie sur $X(\Omega) = \mathbb{N}$ alors :

$$\mathbb{E}(X) = \frac{1-p}{p} \quad \mathbb{V}(X) = \frac{1-p}{p^2} \tag{7.42}$$

Si cette loi est définie sur $X(\Omega) = \mathbb{N}^*$ alors :

$$\mathbb{E}(X) = \frac{1}{p} \quad \mathbb{V}(X) = \frac{1-p}{p^2} \tag{7.43}$$

1.4.3 Autres propriétés

Comme sa loi continue équivalente, *i.e.* la loi exponentielle (▸ section 2.2), la loi géométrique possède la propriété d'être « **sans mémoire** ». L'idée est que lorsque l'on compte le nombre d'échecs avant un succès dans une répétition d'expériences indépendantes de Bernoulli, la probabilité conditionnelle de succès au $k^{ème}$ tirage ne dépend pas du nombre d'échecs préalables. Il n'y a pas de mémoire du nombre d'échecs (ou de succès). Cette propriété est parfois appelée **propriété de Markov**.

Propriété

Absence de mémoire
Si la variable X suit une loi géométrique $\mathcal{G}(p)$ alors pour tout $(t,s) \in \mathbb{R}^{2+}$:

$$\Pr(X > s+t | X > t) = \Pr(X > s) \tag{7.44}$$

Une autre propriété porte sur le lien entre la loi géométrique et la loi binomiale négative. La **loi binomiale négative** ou **loi de Pascal** (du nom du philosophe et mathématicien français Blaise Pascal 1623-1662) correspond à la loi de probabilité de la variable représentant le nombre d'échecs avant l'obtention d'un nombre donné $n \geq 1$ de succès dans une expérience de Bernoulli de paramètre p. Par exemple, on compte le nombre de résultats « face » obtenus avant d'obtenir $n = 3$ fois le résultat positif « pile » (pas forcément consécutivement). Cette loi possède donc deux paramètres comme la loi binomiale, n et p. Mais attention, n désigne dans ce cas un nombre de succès donné et non pas le nombre de répétitions de l'expérience de Bernoulli.

On comprend d'après cette définition qu'une loi géométrique $\mathcal{G}(p)$ définie sur $\mathbb{N}$ n'est rien d'autre qu'une loi binomiale négative de paramètres $n = 1$ et p. De façon générale, on montre le résultat suivant.

Propriété

Somme de lois géométriques indépendantes

Si $X_1, \ldots, X_n$ sont des variables indépendantes distribuées selon une loi géométrique $\mathcal{G}(p)$ alors $\sum_{i=1}^{n} X_i$ suit une loi *binomiale négative* de paramètres n et p.

1.5 Loi de Poisson

La **loi de Poisson**, du nom du mathématicien français Denis Poisson (1781-1840), est une loi de probabilité discrète définie sur l'ensemble des entiers $\mathbb{N}$. Cette loi est notamment utilisée pour représenter un nombre d'événements se produisant dans un laps de temps donné. Dit autrement, c'est une loi permettant de modéliser des **variables de comptage**. Elle est généralement utilisée pour modéliser les **phénomènes d'occurence rare** (► En pratique : une utilisation célèbre de la loi de Poisson) : par exemple le nombre de dépôts de brevets sur une année, le nombre de voitures arrivant à un péage pendant un intervalle de quelques minutes, etc.

La loi de Poisson dépend d'un paramètre réel strictement positif, noté λ, qui comme nous le verrons, correspond à la fois à l'espérance et à la variance de la distribution. Si la variable X suit une loi de Poisson de paramètre $\lambda > 0$ on note :

$$X \sim Pois(\lambda) \quad \text{ou} \quad X \sim \mathcal{P}(\lambda) \tag{7.45}$$

EN PRATIQUE

Une utilisation célèbre de la loi de Poisson

L'exemple le plus célèbre d'utilisation de la loi de Poisson est celui de l'étude de **Ladislaus Bortkiewicz** (économiste et statisticien polonais, 1868-1931) consacrée aux... accidents de mules. Dans un ouvrage de 1908 intitulé *La loi des petits nombres* et consacré à la loi de Poisson, Ladislaus Bortkiewicz applique cette loi de probabilité pour modéliser le nombre d'accidents mortels dans l'armée prussienne dus à des ruades de mules. Rappelons que les armées européennes de la fin du XIX[e] siècle utilisaient plusieurs centaines de milliers de mules ou de chevaux pour le transport des munitions et du ravitaillement. Mais les accidents mortels restaient heureusement bien rares...

1.5.1 Fonction de masse et fonction de répartition

Définition 7.9

La variable aléatoire discrète X définie sur $X(\Omega) = \mathbb{N}$ suit une **loi de Poisson** $\mathcal{P}(\lambda)$ avec $\lambda \in \mathbb{R}^+$, si sa fonction de masse est définie par :

$$f_X(x) = \Pr(X = x) = \frac{\lambda^x \exp(-\lambda)}{x!} \quad \forall x \in \mathbb{N} \tag{7.46}$$

Exemple

Soit Y une variable aléatoire discrète définie sur $Y(\Omega) = \mathbb{N}$ telle que $X \sim \mathcal{P}(0{,}2)$, alors :

$$\Pr(X = 0) = \frac{0{,}2^0 \times \exp(-0{,}2)}{0!} = \exp(-0{,}2) = 0{,}8187 \quad (7.47)$$

$$\Pr(X = 1) = \frac{0{,}2^1 \times \exp(-0{,}2)}{1!} = 0{,}2 \times \exp(-0{,}2) = 0{,}1637 \quad (7.48)$$

$$\Pr(X = 2) = \frac{0{,}2^2 \times \exp(-0{,}2)}{2!} = \frac{0{,}04}{2} \times \exp(-0{,}2) = 0{,}0164 \quad (7.49)$$

Définition 7.10

Si la variable aléatoire X admet une loi $\mathcal{P}(\lambda)$ sur $X(\Omega) = \mathbb{N}$, sa **fonction de répartition** $F_X(x) = \Pr(X \leq x)$ est définie[6] $\forall x \in \mathbb{R}$ par $F_X(x) = 0$, si $x < 0$ et :

$$F_X(x) = \sum_{i=0}^{\lfloor x \rfloor} \Pr(X = i) = \sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^i \exp(-\lambda)}{i!} = \quad \forall x \geq 0 \quad (7.50)$$

où $\lfloor x \rfloor$ désigne le plus grand entier inférieur ou égal à x.

Exemple

Soit Y une variable aléatoire discrète définie sur $Y(\Omega) = \mathbb{N}$ telle que $X \sim \mathcal{P}(0{,}2)$. Calculons les probabilités $\Pr(X \leq 1)$ et $\Pr(X \leq 1{,}27)$.

$$F_X(1) = \Pr(X \leq 1) = \sum_{i=0}^{\lfloor 1 \rfloor} \Pr(X = i) = \sum_{i=0}^{1} \Pr(X = i) \quad (7.51)$$

$$= \Pr(X = 0) + \Pr(X = 1) = 0{,}9824 \quad (7.52)$$

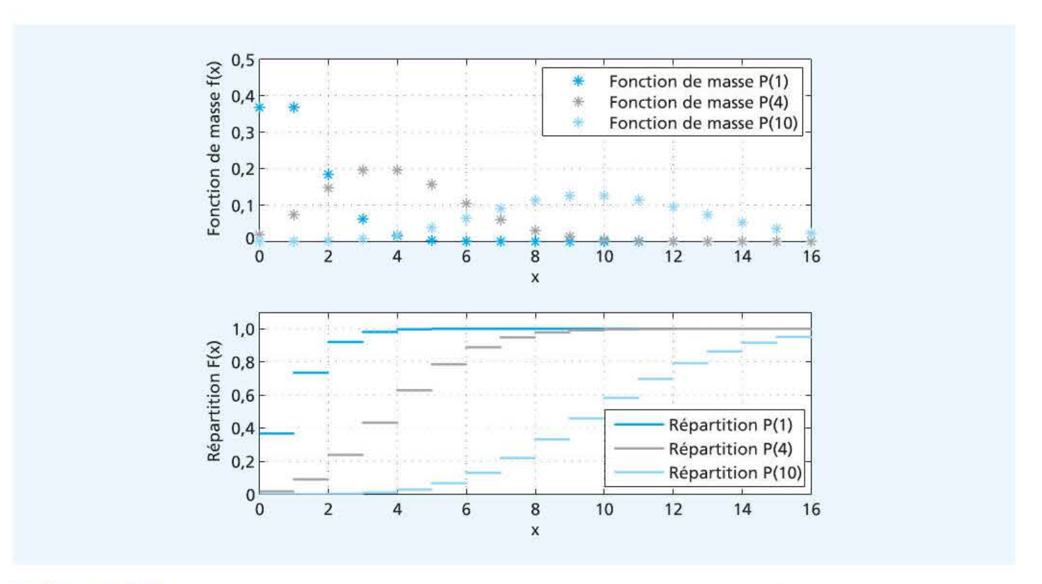
$$F_X(1{,}27) = \Pr(X \leq 1{,}27) = \sum_{i=0}^{\lfloor 1{,}27 \rfloor} \Pr(X = i) = \sum_{i=0}^{1} \Pr(X = i) \quad (7.53)$$

$$= \Pr(X = 0) + \Pr(X = 1) = F_X(1) = 0{,}9824 \quad (7.54)$$

De la même façon que pour la loi binomiale, le calcul des probabilités cumulées $F_X(x)$ dans le cas de la loi de Poisson peut s'avérer fastidieux lorsque x est élevé. On utilise alors soit les fonctions des logiciels de statistique ou des tableurs (par exemple la fonction *LOI.POISSON* sous Excel), soit des tables statistiques de la loi de Poisson. Plusieurs tables figurent en annexe de cet ouvrage pour différentes valeurs du paramètre λ comprises entre 0,1 et 10. Pour chaque λ sont affichées les probabilités cumulées $\Pr(X \leq x)$ obtenues pour différentes valeurs de x allant de 0 à 7 ou de 0 à 30 suivant les cas.

La figure 7.3 représente les fonctions de masse et les fonctions de répartition de trois exemples de lois de Poisson $\mathcal{P}(\lambda)$ pour des paramètres $\lambda = 1$, $\lambda = 4$ et $\lambda = 10$. Comme pour toute variable aléatoire discrète, la fonction de répartition se présente sous la forme d'une fonction en marches d'escalier.

6 Il est aussi possible d'exprimer cette fonction de répartition en fonction de la *fonction gamma incomplète* (► section 2.3).



▲ **Figure 7.3** Exemple de fonctions de masse et de répartition de lois de Poisson

1.5.2 Moments

La fonction génératrice des moments de la loi de Poisson $\mathcal{P}(\lambda)$ est définie par :

$$M_X(t) = \exp(\lambda(\exp(t) - 1)) \quad \forall t \in \mathbb{R} \tag{7.55}$$

De cette fonction génératrice, on peut dériver l'espérance et la variance.

Propriété

Espérance et variance de la loi de Poisson

Si X admet une loi de Poisson de paramètre $\lambda > 0$, alors :

$$\mathbb{E}(X) = \lambda \qquad \mathbb{V}(X) = \lambda \tag{7.56}$$

La particularité de la loi de Poisson est que son espérance est égale à sa variance. Or, dans de nombreux cas pratiques de comptage on observe un **phénomène de sur-dispersion** ce qui signifie que la variance de la variable étudiée est supérieure à son espérance. Par conséquent, la loi de Poisson n'est pas adaptée dans ces cas. On préfère alors utiliser une loi binomiale négative ou une autre loi compatible avec la sur-dispersion.

1.5.3 Autres propriétés

La somme de variables de Poisson est distribuée selon une loi de Poisson.

Propriété

Somme de lois de Poisson

Si les variables $X_1,\ldots,X_n$ sont indépendantes et sont telles que $X_i \sim \mathcal{P}(\lambda_i)$ avec $\lambda_i > 0$ pour $i = 1,\ldots,n$ alors :

$$\sum_{i=1}^{n} X_i \sim \mathcal{P}(\lambda) \quad \text{avec } \lambda = \sum_{i=1}^{n} \lambda_i \tag{7.57}$$

2 Lois usuelles continues

2.1 Loi uniforme continue

La **loi uniforme continue** est une loi de probabilité continue définie sur un intervalle $[a,b] \subset \mathbb{R}$ et caractérisée par une fonction de densité constante pour toutes les valeurs réelles $x \in [a,b]$. Contrairement à la loi uniforme discrète, l'idée d'équiprobabilité de la loi uniforme continue ne se traduit pas au niveau de la probabilité d'une réalisation particulière, puisque pour une loi continue la probabilité d'être en un point est nulle. Elle se traduit par le fait que tous les intervalles de même longueur inclus dans le support $[a,b]$ ont la même probabilité. Si X suit une loi uniforme continue sur le segment $[a,b]$, on note :

$$X \sim \mathcal{U}_{[a,b]} \tag{7.58}$$

2.1.1 Fonction de densité et fonction de répartition

Soient deux valeurs réelles a et b, telles que $b \geq a$.

Définition 7.11

La variable aléatoire réelle X suit une **loi uniforme continue** sur le support $X(\Omega) = [a,b]$ si sa fonction de densité est définie par :

$$f_X(x) = \frac{1}{b-a} \quad \forall x \in X(\Omega) \tag{7.59}$$

Rappelons que $f_X(x) = 0$ si $x \notin X(\Omega) = [a,b]$. Dans le cas particulier où $a = 0$ et $b = 1$, on parle de **loi uniforme (continue) standard**.

Exemple

Considérons une variable aléatoire X distribuée selon une loi uniforme continue sur $X(\Omega) = [0,20]$, sa fonction de densité est définie par $f_X(x) = 1/20$, $\forall x \in [0,20]$ et $f_X(x) = 0$ si $\forall x \notin [0,20]$.

Définition 7.12

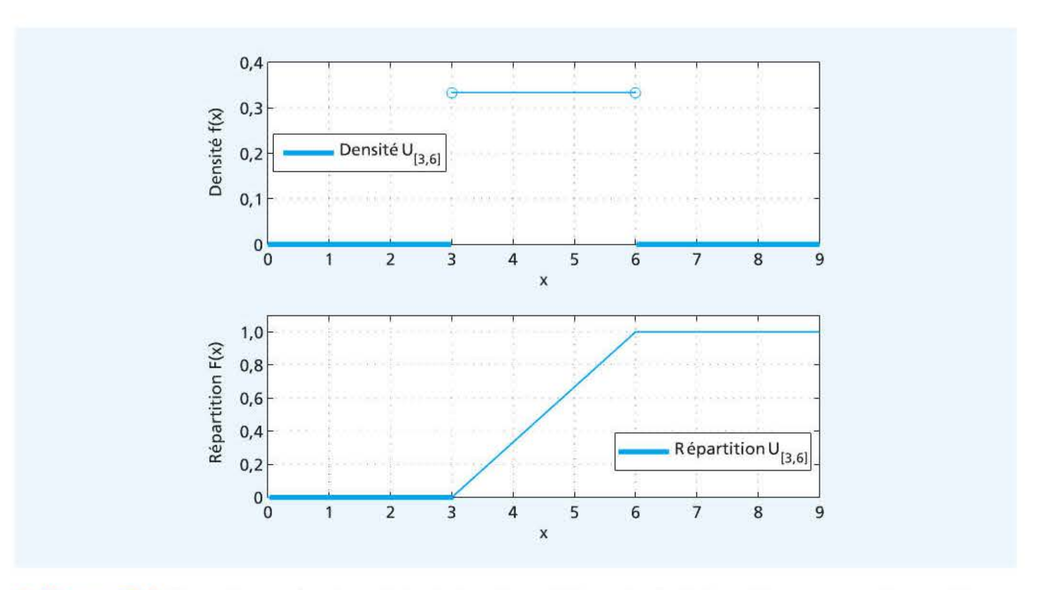
Si la variable aléatoire X admet une loi uniforme continue sur $X(\Omega) = [a,b]$, sa **fonction de répartition** $F_X(x) = \Pr(X \leq x)$ est définie par $\forall x \in \mathbb{R}$:

$$F_X(x) = \begin{cases} 0 & \text{si } x < a \\ \dfrac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } x > b \end{cases} \tag{7.60}$$

Exemple

Soit $X \sim \mathcal{U}_{[0,20]}$, alors $F_X(5) = \Pr(X \leq 5) = (5-0)/20 = 1/4$. Inversement, le fractile d'ordre $\alpha = 0{,}25$ est égal à $F_X^{-1}(0{,}25) = 5$.

Pour illustration, la figure 7.4 représente les fonctions de densité et de répartition de la loi $\mathcal{U}_{[3,6]}$ pour des valeurs de x allant de 0 à 9.



▲ **Figure 7.4 Fonctions de densité et de répartition de la loi uniforme continue $\mathcal{U}_{[3,6]}$**

2.1.2 Moments

La fonction génératrice des moments de la loi uniforme $\mathcal{U}_{[a,b]}$ est égale à :

$$M_X(t) = \frac{\exp(t \times b) - \exp(t \times a)}{t(b-a)} \quad \forall t \in \mathbb{R} \tag{7.61}$$

De cette fonction génératrice, on peut dériver l'espérance et la variance.

Propriété

Espérance et variance de la loi uniforme continue

Si X admet une loi uniforme continue sur $X(\Omega) = [a,b]$, alors :

$$\mathbb{E}(X) = \frac{a+b}{2} \qquad \mathbb{V}(X) = \frac{(b-a)^2}{12} \tag{7.62}$$

Puisque la loi uniforme est **symétrique** par rapport à $\mathbb{E}(X)$, sa skewness est nulle. Sa kurtosis est égale à 1,8 indiquant que cette distribution est **platykurtique** : sa kurtosis est inférieure à celle de la loi normale, égale à 3.

Propriété

Skewness et kurtosis de la loi uniforme

Si $X \sim \mathcal{U}_{[a,b]}$ alors :

$$\text{skewness} = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mathbb{E}(X - \mathbb{E}(X))^3)}{\mathbb{V}(X)^{3/2}} = 0 \tag{7.63}$$

$$\text{kurtosis} = \frac{\mu_4}{\mu_2^2} = \frac{\mathbb{E}(X - \mathbb{E}(X))^4)}{\mathbb{V}(X)^2} = \frac{9}{5} \tag{7.64}$$

2.1.3 Autres propriétés

La loi uniforme standard $\mathcal{U}_{[0,1]}$ est particulièrement utile pour générer des nombres au hasard de n'importe quelle distribution continue (loi normale, loi de Student, etc.). Ce résultat est dû à la propriété dite de transformation intégrale de probabilité ou propriété **PIT** (*probability integral transform*).

Propriété

PIT

Soit Y une variable aléatoire continue admettant une fonction de répartition $F_Y(y)$ définie sur $\mathbb{R}$. Alors, la variable aléatoire $X = F_Y(Y)$ admet une **distribution uniforme standard** :

$$X = F_Y(Y) \sim \mathcal{U}_{[0,1]} \tag{7.65}$$

La propriété PIT est utilisée par tous les logiciels de statistique ou les tableurs pour générer un nombre au hasard y dans n'importe quelle distribution associée à la fonction de répartition $F_Y(y)$. Pour cela, on adopte la démarche suivante :

1. On tire un nombre au hasard dans la loi uniforme standard, c'est-à-dire une réalisation x d'une variable aléatoire X distribuée selon une loi $\mathcal{U}_{[0,1]}$.

2. On cherche la valeur de y telle que $F_Y(y) = x$. Par inversion de la fonction de répartition, il vient $y = F_Y^{-1}(x)$.

Exemple

Soit Y une variable aléatoire réelle telle que $F_Y(y) = 1 - \exp(-y)$. La variable aléatoire $X = F_Y(Y) = 1 - \exp(-Y)$ admet une distribution uniforme sur $[0,1]$. Par conséquent pour tirer un nombre au hasard dans la loi de Y, on commence par tirer un nombre au hasard dans la loi $\mathcal{U}_{[0,1]}$. Si par exemple on obtient une réalisation $x = 0{,}2541$, alors une réalisation y de Y est donnée par :

$$y = F_Y^{-1}(0{,}2541) = -\ln(1 - 0{,}2541) = 0{,}2932 \tag{7.66}$$

2.2 Loi exponentielle

La **loi exponentielle** est une loi de probabilité continue définie sur des valeurs réelles positives. Cette loi correspond au temps mesuré entre des événements issus d'un **processus de Poisson**, *i.e.* un processus continu de comptage dans lequel les événements arrivent de façon continue et indépendamment les uns des autres avec une intensité constante. Tout comme sa loi discrète équivalente (la loi géométrique), une loi exponentielle permet de modéliser la durée de vie d'un phénomène sans mémoire.

La fonction de densité de la loi exponentielle dépend d'un paramètre réel λ strictement positif, appelé **intensité**. Si X suit une loi exponentielle d'intensité $\lambda > 0$ sur $X(\Omega) = \mathbb{R}^+$, on note :

$$X \sim \mathcal{E}xp(\lambda) \tag{7.67}$$

2.2.1 Fonction de densité et fonction de répartition

Définition 7.13

La variable aléatoire réelle X suit une **loi exponentielle** de paramètre $\lambda \in \mathbb{R}^+$ sur le support $X(\Omega) = \mathbb{R}^+$ si sa fonction de densité est définie par :

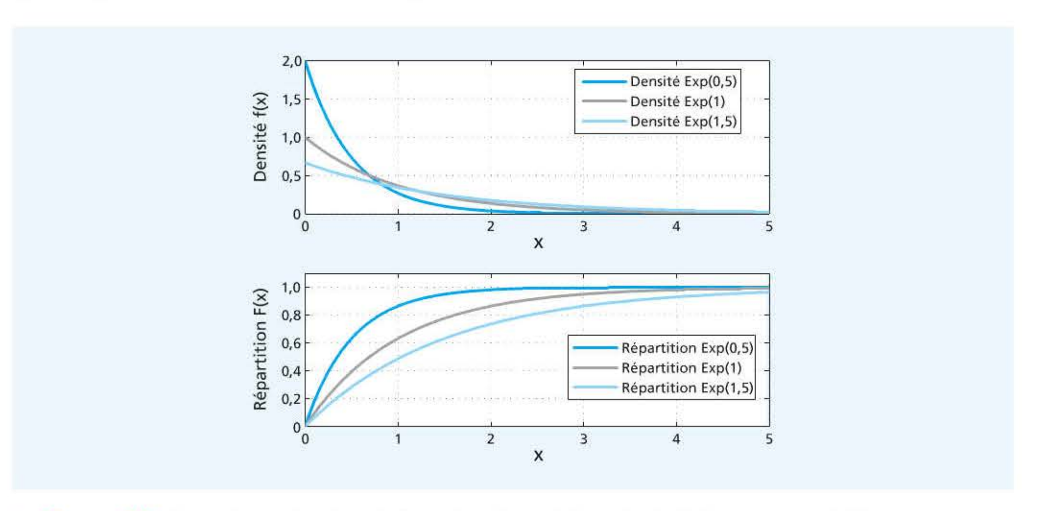
$$f_X(x) = \lambda \exp(-\lambda x) \quad \forall x \in \forall x \in \mathbb{R}^+ \tag{7.68}$$

Définition 7.14

Si la variable aléatoire X admet une loi exponentielle de paramètre $\lambda > 0$ sur $X(\Omega) = \mathbb{R}^+$, sa **fonction de répartition** $F_X(x) = \Pr(X \leq x)$ est définie par :

$$F_X(x) = 1 - \exp(-\lambda x) \quad \forall x \in \mathbb{R} \tag{7.69}$$

La figure 7.5 représente les fonctions de densité et de répartition de trois exemples de lois exponentielles pour des paramètres $\lambda = 0{,}5$, $\lambda = 1$ et $\lambda = 1{,}5$. On vérifie que la fonction de densité de la loi exponentielle est toujours strictement décroissante sur $\mathbb{R}^+$ quelle que soit la valeur de λ (► équation (7.68)).



▲ **Figure 7.5 Fonctions de densité et de répartition de la loi exponentielle**

2.2.2 Moments

La fonction génératrice des moments de la loi $\mathcal{E}xp(\lambda)$ est égale à :

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1} \quad \forall t \in \mathbb{R} \tag{7.70}$$

De cette fonction génératrice, on peut dériver l'espérance et la variance.

Propriété

Espérance et variance de la loi exponentielle

Si X suit une loi exponentielle de paramètre $\lambda > 0$, alors :

$$\mathbb{E}(X) = \frac{1}{\lambda} \qquad \mathbb{V}(X) = \frac{1}{\lambda^2} \tag{7.71}$$

La loi exponentielle n'est pas symétrique par rapport à $\mathbb{E}(X)$, sa skewness est positive ce qui implique que $\Pr(X \geq \mathbb{E}(X)) > \Pr(X \leq \mathbb{E}(X))$. Sa kurtosis est égale à 9 indiquant que cette distribution est **leptokurtique** (▶ chapitre 6).

Propriété

Skewness et kurtosis de la loi exponentielle

Si $X \sim \mathcal{E}xp(\lambda)$ alors :

$$\text{skewness} = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mathbb{E}\left(X - \mathbb{E}(X)\right)^3)}{\mathbb{V}(X)^{3/2}} = 2 \tag{7.72}$$

$$\text{kurtosis} = \frac{\mu_4}{\mu_2^{2}} = \frac{\mathbb{E}\left(X - \mathbb{E}(X)\right)^4)}{\mathbb{V}(X)^{2}} = 9 \tag{7.73}$$

2.2.3 Autres propriétés

Tout comme la loi géométrique, la loi exponentielle est une loi sans mémoire. Elle satisfait la **propriété de Markov** (▶ section 1.4).

2.3 Loi normale

La **loi normale**, ou loi de Laplace-Gauss[7], est une loi de probabilité continue définie sur l'ensemble des réels $\mathbb{R}$. C'est sans conteste la loi de probabilité continue la plus utilisée, notamment en raison du théorème central limite que nous étudierons dans le chapitre 8. La densité de la loi normale dépend d'un paramètre de location (qui correspond à son **espérance**) noté[8] μ et d'un paramètre d'échelle (qui correspond à sa **variance**) noté σ^2. Si une variable aléatoire X définie sur $X(\Omega) = \mathbb{R}$, suit une loi normale de paramètres μ et σ^2, on note :

$$X \sim \mathcal{N}\left(\mu,\sigma^2\right) \tag{7.74}$$

Remarque : La loi normale est parfois notée sous la forme $X \sim \mathcal{N}(\mu,\sigma)$ où σ désigne l'écart-type de la distribution de X. Il convient de ne pas confondre les deux notations.

2.3.1 Fonction de densité et fonction de répartition

Définition 7.15

La variable aléatoire réelle X suit une **loi normale** d'espérance μ et de variance σ^2, notée $\mathcal{N}\left(\mu,\sigma^2\right)$, si sa fonction de densité est définie sur $X(\Omega) = \mathbb{R}$ par :

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \quad \forall x \in \mathbb{R} \tag{7.75}$$

avec $\mu \in \mathbb{R}$ et $\sigma \in \mathbb{R}^+$.

Le symbole π renvoie au nombre π égal à approximativement 3,1415.

7 Du nom du mathématicien français Pierre-Simon Laplace (1749-1827) et du mathématicien allemand Carl Friedrich Gauss (1777-1855).

8 Il ne faut pas confondre la notation μ de l'espérance (moment ordinaire d'ordre 1, *i.e.* m_1) utilisée pour la loi normale et la notation des moments centrés d'ordre k, notés μ_k.

Propriété

Loi normale

Si la variable aléatoire réelle X suit une loi normale $\mathcal{N}(\mu,\sigma^2)$ alors sa fonction de densité vérifie les propriétés suivantes :

1. $\lim_{x\to+\infty} f_X(x) = \lim_{x\to-\infty} f_X(x) = 0.$
2. $f_X(\mu + x) = f_X(\mu - x), \forall x \in \mathbb{R}.$
3. $f_X(x)$ atteint son maximum en $x = \mu$.

La première propriété n'est pas propre à la loi normale. La seconde propriété signifie que la fonction de densité de la loi normale est symétrique par rapport à son espérance μ. La troisième propriété implique que le **mode** de la distribution normale est égal à son espérance. Comme le montre la figure 7.7, la distribution normale est **unimodale**, c'est-à-dire qu'elle ne possède qu'un seul mode (▶ chapitre 1). On rappelle enfin que comme pour toute fonction de densité, la densité de la loi normale intègre à 1 sur son support, *i.e.* $\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$.

Parmi les lois normales générales, on distingue la **loi normale centrée réduite** ou **loi normale standard**, d'espérance nulle et de variance égale à 1. On admet (cf. propriétés) que :

$$X \sim \mathcal{N}(\mu,\sigma^2) \iff \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1) \tag{7.76}$$

Par convention, la fonction de densité de la loi normale centrée réduite $\mathcal{N}(0,1)$ est notée $\phi(x)$ (prononcer « *phi de x* »).

Définition 7.16

La variable aléatoire réelle X suit une **loi normale centrée réduite** $\mathcal{N}(0,1)$ si sa fonction de densité est définie par :

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad \forall x \in \mathbb{R} \tag{7.77}$$

On peut toujours exprimer la fonction de densité d'une loi $\mathcal{N}(\mu,\sigma^2)$ en fonction de la densité de la loi normale centrée réduite. En effet, si l'on note $f_{\mu,\sigma^2}(x)$ la densité de la loi $\mathcal{N}(\mu,\sigma^2)$ il vient :

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) \quad \forall x \in \mathbb{R} \tag{7.78}$$

Définition 7.17

Si la variable aléatoire réelle X admet une loi normale $\mathcal{N}(\mu,\sigma^2)$, sa **fonction de répartition** $F_X(x) = \Pr(X \leq x)$ est définie par $\forall x \in \mathbb{R}$:

$$F_X(x) = \int_{-\infty}^{x} f_X(z)\,dz = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right) dz \tag{7.79}$$

Cette fonction de répartition n'a pas d'**expression analytique** (▶ chapitre 6).

Le fait que la fonction de répartition n'ait pas d'expression analytique signifie qu'elle ne s'exprime pas à partir de fonctions usuelles (log, exponentielle, etc.) mais qu'elle devient elle-même une fonction usuelle. Par conséquent, si l'on souhaite calculer une probabilité cumulée pour une loi normale on doit nécessairement recourir à une table statistique ou à un logiciel de statistique (par exemple la fonction *LOI.NORMALE* sous Excel). La table statistique fournie en annexe (▸ figure 7.6 pour un extrait) se réfère uniquement au cas particulier de la loi normale centrée réduite $\mathcal{N}(0,1)$. En effet, il est toujours possible d'exprimer les probabilités cumulées d'une loi normale générale $\mathcal{N}(\mu,\sigma^2)$ en fonction de celles de la loi $\mathcal{N}(0,1)$.

Définition 7.18

La **fonction de répartition** de la loi normale centrée réduite $\mathcal{N}(0,1)$, notée $\Phi(x)$ (prononcer « *grand phi de x* »), est définie par $\forall x \in \mathbb{R}$:

$$\Phi(x) = \int_{-\infty}^{x} \phi(z)\,dz = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \qquad (7.80)$$

Supposons que $X \sim \mathcal{N}(\mu,\sigma^2)$ et que l'on veuille calculer $F_X(c) = \Pr(X \leq c)$ où c est une valeur réelle. On peut alors exprimer cette probabilité cumulée à l'aide de la fonction de répartition $\Phi(.)$ de la loi $\mathcal{N}(0,1)$.

$$F_X(c) = \Pr(X \leq c) = \Pr\left(\frac{X-\mu}{\sigma} \leq \frac{c-\mu}{\sigma}\right) = \Phi\left(\frac{c-\mu}{\sigma}\right) \qquad (7.81)$$

puisque la variable centrée réduite $(X-\mu)/\sigma$ suit une loi normale centrée réduite.

Exemple

Soit X une variable aléatoire réelle telle que $X \sim \mathcal{N}(0{,}6\ ;\ 4)$. Calculons la probabilité cumulée $\Pr(X \leq -0{,}5)$.

$$F_X(-0{,}5) = \Pr(X \leq -0{,}5) = \Pr\left(\frac{X-0{,}6}{\sqrt{4}} \leq \frac{-0{,}5-0{,}6}{\sqrt{4}}\right) = \Phi(-0{,}55) \qquad (7.82)$$

Si l'on admet que $\Phi(-0{,}55) = 0{,}2912$ on obtient $\Pr(X \leq -0{,}5) = 0{,}2912$.

Comment calculer une probabilité cumulée pour une loi normale centrée réduite à partir de la table statistique de la figure 7.6 ? Cette table permet de déterminer la probabilité cumulée $\Phi(z)$ pour un certain nombre de valeurs réelles z. La valeur de z est reconstruite par addition des valeurs reportées en lignes (allant de 0 à 0,09 par pas de 0,01) et des valeurs reportées en colonne (allant de 0 à 2,9 par pas de 0,1). Par exemple, si l'on souhaite calculer la probabilité cumulée $\Phi(1{,}02)$ on décompose la valeur 1,02 en une somme 1 + 0,02. En considérant la ligne correspondant à la valeur 1 et la colonne correspondant à la valeur 0,02, on trouve à l'intersection la valeur de la probabilité cumulée $\Phi(1{,}02) = 0{,}846136$. Pour des valeurs de z allant de 3 à 4,5, un tableau spécifique (non reproduit sur la figure 7.6) permet de lire directement la valeur de la fonction de répartition.

On constate que les réalisations z reportées dans la table de la figure 7.6 sont toutes positives. Dès lors, comment calculer une probabilité cumulée du type $\Phi(-0{,}7)$? On utilise pour cela la propriété suivante.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,500000	0,503989	0,507978	0,511966	0,515953	0,519939	0,523922	0,527903	0,531881	0,535856
0,1	0,539828	0,543795	0,547758	0,551717	0,555670	0,559618	0,563559	0,567495	0,571424	0,575345
0,2	0,579260	0,583166	0,587064	0,590954	0,594835	0,598706	0,602568	0,606420	0,610261	0,614092
0,3	0,617911	0,621720	0,625516	0,629300	0,633072	0,636831	0,640576	0,644309	0,648027	0,651732
0,4	0,655422	0,659097	0,662757	0,666402	0,670031	0,673645	0,677242	0,680822	0,684386	0,687933
0,5	0,691462	0,694974	0,698468	0,701944	0,705401	0,708840	0,712260	0,715661	0,719043	0,722405
0,6	0,725747	0,729069	0,732371	0,735653	0,738914	0,742154	0,745373	0,748571	0,751748	0,754903
0,7	0,758036	0,761148	0,764238	0,767305	0,770350	0,773373	0,776373	0,779350	0,782305	0,785236
0,8	0,788145	0,791030	0,793892	0,796731	0,799546	0,802337	0,805105	0,807850	0,810570	0,813267
0,9	0,815940	0,818589	0,821214	0,823814	0,826391	0,828944	0,831472	0,833977	0,836457	0,838913
1,0	0,841345	0,843752	0,846136	0,848495	0,850830	0,853141	0,855428	0,857690	0,859929	0,862143
1,1	0,864334	0,866500	0,868643	0,870762	0,872857	0,874928	0,876976	0,879000	0,881000	0,882977
1,2	0,884930	0,886861	0,888768	0,890651	0,892512	0,894350	0,896165	0,897958	0,899727	0,901475
1,3	0,903200	0,904902	0,906582	0,908241	0,909877	0,911492	0,913085	0,914657	0,916207	0,917736
1,4	0,919243	0,920730	0,922196	0,923641	0,925066	0,926471	0,927855	0,929219	0,930563	0,931888
1,5	0,933193	0,934478	0,935745	0,936992	0,938220	0,939429	0,940620	0,941792	0,942947	0,944083
1,6	0,945201	0,946301	0,947384	0,948449	0,949497	0,950529	0,951543	0,952540	0,953521	0,954486
1,7	0,955435	0,956367	0,957284	0,958185	0,959070	0,959941	0,960796	0,961636	0,962462	0,963273
1,8	0,964070	0,964852	0,965620	0,966375	0,967116	0,967843	0,968557	0,969258	0,969946	0,970621
1,9	0,971283	0,971933	0,972571	0,973197	0,973810	0,974412	0,975002	0,975581	0,976148	0,976705
2,0	0,977250	0,977784	0,978308	0,978822	0,979325	0,979818	0,980301	0,980774	0,981237	0,981691
2,1	0,982136	0,982571	0,982997	0,983414	0,983823	0,984222	0,984614	0,984997	0,985371	0,985738
2,2	0,986097	0,986447	0,986791	0,987126	0,987455	0,987776	0,988089	0,988396	0,988696	0,988989
2,3	0,989276	0,989556	0,989830	0,990097	0,990358	0,990613	0,990863	0,991106	0,991344	0,991576
2,4	0,991802	0,992024	0,992240	0,992451	0,992656	0,992857	0,993053	0,993244	0,993431	0,993613
2,5	0,993790	0,993963	0,994132	0,994297	0,994457	0,994614	0,994766	0,994915	0,995060	0,995201
2,6	0,995339	0,995473	0,995604	0,995731	0,995855	0,995975	0,996093	0,996207	0,996319	0,996427
2,7	0,996533	0,996636	0,996736	0,996833	0,996928	0,997020	0,997110	0,997197	0,997282	0,997365
2,8	0,997445	0,997523	0,997599	0,997673	0,997744	0,997814	0,997882	0,997948	0,998012	0,998074
2,9	0,998134	0,998193	0,998250	0,998305	0,998359	0,998411	0,998462	0,998511	0,998559	0,998605

▲ **Figure 7.6** Extrait de la table de la loi normale centrée réduite

Propriété

Fonction de répartition de la loi normale centrée réduite

Puisque la densité de la loi normale centrée réduite $\mathcal{N}(0,1)$ est symétrique par rapport à son espérance égale à 0, on a :

$$\Phi(0) = 0{,}5 \tag{7.83}$$

$$\Phi(-x) = 1 - \Phi(x) \quad \forall x \in \mathbb{R} \tag{7.84}$$

Exemple

On suppose que $X \sim \mathcal{N}(0{,}6\,;\,4)$, déterminons les probabilités $\Pr(X \geq 1{,}86)$ et $\Pr(X \leq -0{,}22)$ à partir de la table statistique de la loi normale centrée réduite.

$$\Pr(X \geq 1{,}86) = 1 - \Pr(X \leq 1{,}86) \tag{7.85}$$

$$= 1 - \Pr\left(\frac{X - 0{,}6}{\sqrt{4}} \leq \frac{1{,}86 - 0{,}6}{\sqrt{4}}\right) \tag{7.86}$$

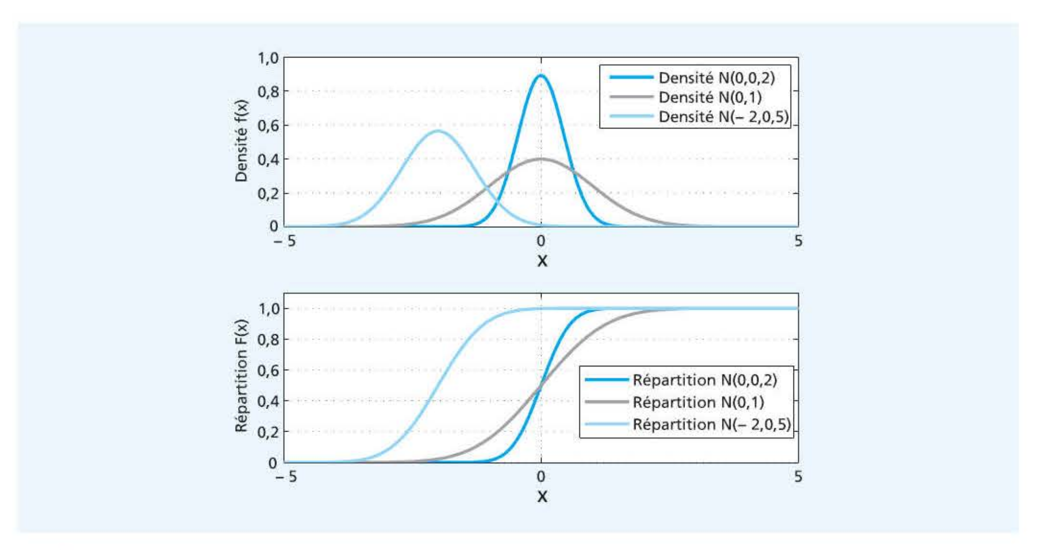
$$= 1 - \Phi(0{,}63) = 1 - 0{,}735653 = 0{,}264347 \tag{7.87}$$

$$\Pr(X \leq -0{,}22) = \Pr\left(\frac{X - 0{,}6}{\sqrt{4}} \leq \frac{-0{,}22 - 0{,}6}{\sqrt{4}}\right) \tag{7.88}$$

$$= \Phi(-0{,}41) = 1 - \Phi(0{,}41) \tag{7.89}$$

$$= 1 - 0{,}659097 = 0{,}340903 \tag{7.90}$$

La figure 7.7 représente les fonctions de densité et de répartition de trois exemples de lois normales $\mathcal{N}\left(\mu,\sigma^2\right)$ pour les couples $\left(\mu,\sigma^2\right) = (0\ ;\ 0{,}2)$, $(0{,}1)$ et $(-2\ ;\ 0{,}5)$. On observe que lorsque l'espérance μ varie, cela déplace la fonction de densité vers la gauche ou vers la droite sans changer sa forme. C'est pourquoi le paramètre μ est un **paramètre de location**. En revanche, le fait de modifier la variance σ^2 déforme l'allure de la densité : plus la variance est importante, plus la fonction de densité s'aplatit. On dit que la variance est un **paramètre d'échelle**.



▲ **Figure 7.7 Exemple de fonctions de densité et de répartition de lois normales**

2.3.2 Quantiles de la loi normale

Les **quantiles de la loi normale** s'obtiennent par inversion de la fonction de répartition (► chapitre 6). Rappelons que le quantile d'ordre $\alpha \in [0,1]$, noté Q_α ou $F_X^{-1}(\alpha)$, est la réalisation telle que $\Pr\left(X \leq F_X^{-1}(\alpha)\right) = F_X(F_X^{-1}(\alpha)) = \alpha$. Notons que l'on peut *toujours* obtenir les quantiles d'une loi normale $\mathcal{N}\left(\mu,\sigma^2\right)$ à partir des quantiles d'une loi normale centrée réduite, notés $\Phi^{-1}(\alpha)$. En effet, si l'on suppose que $X \sim \mathcal{N}\left(\mu,\sigma^2\right)$ et que l'on note $F_X^{-1}(\alpha)$ le quantile d'ordre α de cette loi, par définition :

$$\Pr\left(X \leq F_X^{-1}(\alpha)\right) = \Pr\left(\frac{X-\mu}{\sigma} \leq \frac{F_X^{-1}(\alpha)-\mu}{\sigma}\right) = \alpha \tag{7.91}$$

Sachant que $(X-\mu)/\sigma$ suit une loi normale centrée réduite, il vient :

$$\Phi\left(\frac{F_X^{-1}(\alpha)-\mu}{\sigma}\right) = \alpha \tag{7.92}$$

En appliquant la fonction $\Phi^{-1}(.)$ aux deux membres de cette égalité, il vient :

$$\Phi^{-1}\left(\Phi\left(\frac{F_X^{-1}(\alpha)-\mu}{\sigma}\right)\right) = \Phi^{-1}(\alpha) \tag{7.93}$$

ou encore :

$$\frac{F_X^{-1}(\alpha)-\mu}{\sigma} = \Phi^{-1}(\alpha) \tag{7.94}$$

Propriété

Quantiles

On peut toujours exprimer le quantile $F_X^{-1}(\alpha)$ d'une loi $\mathcal{N}\left(\mu,\sigma^2\right)$ en fonction du quantile $\Phi^{-1}(\alpha)$ de la loi normale centrée réduite comme :

$$F_X^{-1}(\alpha) = \mu + \sigma\Phi^{-1}(\alpha) \quad \forall\alpha \in [0,1] \tag{7.95}$$

Exemple

Soit $X \sim \mathcal{N}(0{,}6\ ;\ 4)$. Déterminons le quantile d'ordre $\alpha = 10\ \%$ de cette loi en supposant que $\Phi^{-1}(0{,}1) = -1{,}2816$. D'après la relation de l'équation (7.95), il vient :

$$F_X^{-1}(0{,}1) = 0{,}6 + \sqrt{4} \times \Phi(0{,}1) = 0{,}6 + \sqrt{4} \times (-1{,}2816) = -1{,}9632$$

Par définition si $X \sim \mathcal{N}(0{,}6\ ;\ 4)$ alors $\Pr(X \leq -1{,}9632) = 0{,}1$.

Comment déterminer les quantiles de la loi normale centrée réduite ? Puisque la fonction de répartition de cette loi n'a pas de forme analytique, sa fonction de répartition inverse $\Phi^{-1}(\alpha)$ n'en a pas non plus. On doit donc recourir à des tables statistiques ou à des logiciels (par exemple la fonction *LOI.NORMALE.INVERSE* d'Excel). Il est possible d'utiliser la *table de la fonction de répartition*[9] de la loi normale centrée réduite (► figure 7.6) pour retrouver les fractiles $\Phi^{-1}(\alpha)$. Tout d'abord remarquons que les valeurs de cette table correspondent à des probabilités cumulées qui sont toutes supérieures à 0,5, puisque la fonction de répartition est évaluée pour des valeurs $z \geq 0$ pour lesquelles $\Phi(z) \geq 0{,}5$. Dès lors, deux cas de figure doivent être distingués suivant la valeur de la probabilité α.

Premier cas. Si l'on cherche un quantile pour une probabilité $\alpha \geq 0{,}5$, la lecture de ce quantile se fait directement. On cherche dans la table la valeur la plus proche de α (ou les valeurs encadrant α) et l'on reconstruit la valeur du quantile $\Phi^{-1}(\alpha)$ par addition des nombres figurant en en-tête de colonne et de ligne. Par exemple, si l'on cherche $\Phi^{-1}(0{,}95)$, les valeurs les plus proches de 0,95 figurant dans la table sont 0,949497 et 0,950529. Ces valeurs correspondent respectivement à des réalisations égales à 1,64 $(1{,}6 + 0{,}04)$ et 1,65 $(1 + 0{,}05)$. Le quantile à 95 % est donc compris entre 1,64 et 1,65.

Deuxième cas. Si l'on cherche un quantile pour une probabilité $\alpha < 0{,}5$, on utilise la propriété suivante.

Propriété

Quantiles de la loi normale centrée réduite

Puisque la densité de loi normale centrée réduite $\mathcal{N}(0,1)$ est symétrique par rapport à son espérance égale à 0, sa fonction de répartition inverse vérifie :

$$\Phi^{-1}(0{,}5) = 0$$

$$\Phi^{-1}(\alpha) = -\Phi^{-1}(1-\alpha) \quad \forall\alpha \in [0,1] \tag{7.96}$$

9 Pour une mesure plus précise des quantiles, on peut aussi utiliser des « tables de quantiles » de la loi normale centrée réduite.

Exemple

Déterminons le quantile d'ordre $\alpha = 0{,}025$ de la loi normale centrée réduite. Par définition $\Phi^{-1}(0{,}025) = -\Phi^{-1}(0{,}975)$. On cherche alors le quantile à 97,5 % dans la table de la figure 7.6 selon la méthode précédente. On trouve que ce quantile est compris entre 1,95 et 1,96. Par conséquent, $\Phi^{-1}(0{,}025)$ est compris entre $-1{,}95$ et $-1{,}96$.

Appliquons à présent cette démarche pour déterminer les quantiles d'une loi normale générale $\mathcal{N}\left(\mu,\sigma^2\right)$.

Exemple

On suppose que $X \sim \mathcal{N}(0{,}6\ ;\ 4)$, déterminons les quantiles $F_X^{-1}(\alpha)$ d'ordres $\alpha = 0{,}05$ et $\alpha = 0{,}90$. Nous savons que :

$$F_X^{-1}(0{,}05) = 0{,}6 + \sqrt{4} \times \Phi^{-1}(0{,}05) \tag{7.97}$$

Sachant que $\Phi^{-1}(0{,}05) = -\Phi^{-1}(0{,}95)$, on cherche dans la table de la figure 7.6 la valeur la plus proche de 0,95. On trouve la probabilité 0,950529 associée à une réalisation de 1,65 $(1{,}6 + 0{,}05)$. Par conséquent $\Phi^{-1}(0{,}05) \simeq -1{,}65$.

$$F_X^{-1}(0{,}05) = 0{,}6 + 2 \times (-1{,}65) \simeq -2{,}70 \tag{7.98}$$

On obtient le quantile $F_X^{-1}(0{,}90)$ par lecture directe de la table :

$$F_X^{-1}(0{,}90) = 0{,}6 + \sqrt{4} \times \Phi^{-1}(0{,}90) \simeq 0{,}6 + 2 \times 1{,}28 \simeq 3{,}16 \tag{7.99}$$

2.3.3 Moments

La fonction génératrice des moments de la loi normale $\mathcal{N}\left(\mu,\sigma^2\right)$ est égale à :

$$M_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) \quad \forall t \in \mathbb{R} \tag{7.100}$$

Par définition, l'espérance et la variance sont égales aux paramètres de la loi :

$$\mathbb{E}(X) = \mu \qquad \mathbb{V}(X) = \sigma^2 \tag{7.101}$$

La loi normale est une distribution *symétrique* par rapport à $\mathbb{E}(X)$ et *mesokurtique.*

Propriété

Skewness et kurtosis de la loi normale

Si $X \sim \mathcal{N}\left(\mu,\sigma^2\right)$ alors :

$$\text{Skewness} = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mathbb{E}\left(X - \mathbb{E}(X)\right)^3)}{\mathbb{V}(X)^{3/2}} = 0 \tag{7.102}$$

$$\text{Kurtosis} = \frac{\mu_4}{\mu_2^{2}} = \frac{\mathbb{E}\left(X - \mathbb{E}(X)\right)^4)}{\mathbb{V}(X)^{2}} = 3 \tag{7.103}$$

2.3.4 Autres propriétés

Propriété

Linéarité des lois normales

Soit une variable aléatoire réelle $X \sim \mathcal{N}\left(\mu,\sigma^2\right)$ et soient deux constantes $(a,b) \in \mathbb{R}^2$, alors :

$$a + bX \sim \mathcal{N}\left(a + b\mu, b^2\sigma^2\right) \qquad (7.104)$$

Ainsi, la transformée linéaire d'une variable normalement distribuée suit une loi normale. Puisque l'espérance est un opérateur linéaire et la variance est un opérateur quadratique (▶ chapitre 6), les moments de cette loi sont définis par :

$$\mathbb{E}(a + bX) = a + b\mathbb{E}(X) = a + b\mu \qquad (7.105)$$

$$\mathbb{V}(a + bX) = b^2\mathbb{V}(X) = b^2\sigma^2 \qquad (7.106)$$

C'est ce résultat qui explique notamment que :

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0,1) \Longleftrightarrow X \sim \mathcal{N}\left(\mu,\sigma^2\right) \qquad (7.107)$$

En effet, pour retrouver ce résultat il suffit d'appliquer la transformation linéaire $\mu+\sigma Z$ à la variable normale centrée réduite Z pour montrer que $X \sim \mathcal{N}\left(\mu,\sigma^2\right)$.

Propriété

Somme de variables normales indépendantes

Soient $X_1,\ldots,X_n$ des variables aléatoires réelles indépendantes telles que $X_i \sim \mathcal{N}\left(\mu_i,\sigma_i^2\right)$ pour $i = 1,\ldots,n$. Alors :

$$\sum_{i=1}^{n} X_i^2 \sim \mathcal{N}\left(\mu,\sigma^2\right) \qquad (7.108)$$

avec $\mu = \sum_{i=1}^{n} \mu_i$ et $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$.

2.4 Loi du khi-deux

La **loi du khi-deux** (ou « khi carré ») est une loi de probabilité continue définie sur l'ensemble des réels positifs $\mathbb{R}^+$. Sa densité dépend d'un paramètre appelé nombre de **degrés de liberté**, noté k, avec $k \in \mathbb{N}^*$. Si une variable aléatoire X définie sur $X(\Omega) = \mathbb{R}^+$ suit une loi du khi-deux à k degrés de liberté, on note alors :

$$X \sim \chi^2(k) \qquad (7.109)$$

Remarque : Le nombre de dégrés de liberté k est un entier non nul, on ne peut donc pas définir une loi du khi-deux $\chi^2(0)$ ou $\chi^2(1/2)$ par exemple.

La distribution du khi-deux à k degrés de liberté correspond à la distribution de la somme des carrés de k variables aléatoires indépendantes admettant une distribution normale centrée et réduite.

Propriété

Définition d'une loi de khi-deux

Soient $X_1,\dots,X_k$ des variables aléatoires réelles indépendantes telles que $X_i \sim \mathcal{N}(0,1)$ pour $i = 1,\dots,k$. Alors :

$$\sum_{i=1}^{k} X_i \sim \chi^2(k) \tag{7.110}$$

Par définition, si la variable X suit une loi normale centrée réduite, la variable X^2 suit une distribution du khi-deux à 1 degré de liberté.

Remarque : Puisque la loi du khi-deux correspond à la loi de la somme de *carrés* de variables normales, cette loi ne peut être définie que sur $\mathbb{R}^+$: la *réalisation d'une variable du khi-deux n'est jamais négative*.

2.4.1 Fonction de densité et fonction de répartition

Définition 7.19

La variable aléatoire réelle X suit une **loi du khi-deux** à $k \in \mathbb{N}^*$ degrés de liberté sur le support $X(\Omega) = \mathbb{R}^+$ si sa fonction de densité est définie par :

$$f_X(x) = \frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} \exp\left(-\frac{x}{2}\right) \quad \forall x \in X(\Omega) \tag{7.111}$$

Cette densité fait apparaître la fonction gamma[10], notée $\Gamma(.)$ (prononcer « grand gamma »).

Définition 7.20

La fonction gamma, notée $\Gamma(z)$, est définie pour tout entier $z \in \mathbb{R}$ par :

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} \exp(-t)\, dt \tag{7.112}$$

Cette fonction gamma est programmée dans tous les logiciels de statistique et les tableurs (fonction *GAMMA* par exemple sous Excel). Elle possède certaines propriétés qui simplifient souvent les calculs.

Propriété

Fonction gamma

La fonction gamma vérifie :

1. Si z est un entier, alors $\Gamma(z) = (z-1)!$
2. $\forall z \in \mathbb{R}, \Gamma(z) = (z-1)\,\Gamma(z-1)$.
3. $\Gamma(1/2) = \sqrt{\pi}$.

10 Il convient de ne pas confondre la *fonction* gamma et la *loi* gamma. Il existe en effet une loi de probabilité gamma dont la densité dépend de deux paramètres α et β.

Exemple

Soit $X \sim \chi^2(4)$. La densité de X évaluée au point $x = 2$ est égale à :

$$f_X(2) = \frac{1}{2^{\frac{4}{2}} \times \Gamma\left(\frac{4}{2}\right)} \times 2^{\frac{4}{2}-1} \times \exp\left(-\frac{2}{2}\right) \tag{7.113}$$

Sachant que $\Gamma(2) = (2-1)! = 1$, on obtient :

$$f_X(2) = \frac{1}{2^2 \times 1} \times 2 \times \exp(-1) = \frac{\exp(-1)}{2} = 0{,}1839 \tag{7.114}$$

Comme pour la loi normale, il n'existe pas de forme analytique pour la fonction de répartition de la loi du khi-deux. Par conséquent si l'on souhaite calculer une *probabilité cumulée* pour une loi du khi-deux on doit recourir à une table statistique, à un logiciel de statistique ou à un tableur (par exemple la fonction *LOI.KHIDEUX* sous Excel). Une table de la fonction de répartition de la loi du khi-deux est fournie en annexe (► figure 7.8 pour un extrait). Sur cette table sont reportées les valeurs z ayant une probabilité p d'être dépassées, *i.e.* $\Pr(X \geq z) = p$. Les probabilités p figurent en en-têtes de colonnes et vont de 0,001 à 0,99. En ligne figure le nombre de degrés de liberté k de la loi du khi-deux qui varie de 1 à 120 (non reproduit intégralement sur la figure 7.8).

r	P= 0,990	P= 0,975	P= 0,950	P= 0,900	P= 0,800	P= 0,700	P= 0,500	P= 0,300	P= 0,200	P= 0,100	P= 0,010	P= 0,005	P= 0,001
1	0,000	0,001	0,004	0,016	0,064	0,148	0,455	1,074	1,642	2,706	6,635	7,879	10,828
2	0,020	0,051	0,103	0,211	0,446	0,713	1,386	2,408	3,219	4,605	9,210	10,597	13,816
3	0,115	0,216	0,352	0,584	1,005	1,424	2,366	3,665	4,642	6,251	11,345	12,838	16,266
4	0,297	0,484	0,711	1,064	1,649	2,195	3,357	4,878	5,989	7,779	13,277	14,860	18,467
5	0,554	0,831	1,145	1,610	2,343	3,000	4,351	6,064	7,289	9,236	15,086	16,750	20,515
6	0,872	1,237	1,635	2,204	3,070	3,828	5,348	7,231	8,558	10,645	16,812	18,548	22,458
7	1,239	1,690	2,167	2,833	3,822	4,671	6,346	8,383	9,803	12,017	18,475	20,278	24,322
8	1,646	2,180	2,733	3,490	4,594	5,527	7,344	9,524	11,030	13,362	20,090	21,955	26,124
9	2,088	2,700	3,325	4,168	5,380	6,393	8,343	10,656	12,242	14,684	21,666	23,589	27,877
10	2,558	3,247	3,940	4,865	6,179	7,267	9,342	11,781	13,442	15,987	23,209	25,188	29,588
11	3,053	3,816	4,575	5,578	6,989	8,148	10,341	12,899	14,631	17,275	24,725	26,757	31,264
12	3,571	4,404	5,226	6,304	7,807	9,034	11,340	14,011	15,812	18,549	26,217	28,300	32,909
13	4,107	5,009	5,892	7,042	8,634	9,926	12,340	15,119	16,985	19,812	27,688	29,819	34,528
14	4,660	5,629	6,571	7,790	9,467	10,821	13,339	16,222	18,151	21,064	29,141	31,319	36,123
15	5,229	6,262	7,261	8,547	10,307	11,721	14,339	17,322	19,311	22,307	30,578	32,801	37,697
16	5,812	6,908	7,962	9,312	11,152	12,624	15,338	18,418	20,465	23,542	32,000	34,267	39,252
17	6,408	7,564	8,672	10,085	12,002	13,531	16,338	19,511	21,615	24,769	33,409	35,718	40,790
18	7,015	8,231	9,390	10,865	12,857	14,440	17,338	20,601	22,760	25,989	34,805	37,156	42,312
19	7,633	8,907	10,117	11,651	13,716	15,352	18,338	21,689	23,900	27,204	36,191	38,582	43,820
20	8,260	9,591	10,851	12,443	14,578	16,266	19,337	22,775	25,038	28,412	37,566	39,997	45,315

▲ **Figure 7.8 Extrait de la table de la loi du khi-deux**

Exemple

Supposons que $X \sim \chi^2(5)$ et que l'on souhaite calculer la probabilité cumulée $F_X(1{,}61) = \Pr(X \leq 1{,}61)$. Par définition, on a $\Pr(X \leq 1{,}61) = 1 - \Pr(X \geq 1{,}61)$. Sur la ligne $k = 5$ on cherche la valeur la plus proche de 1,61. Cette valeur figure dans le tableau et correspond à une probabilité p égale à 0,9. Par conséquent, $F_X(1{,}61) = \Pr(X \leq 1{,}61) = 1 - 0{,}9 = 0{,}10$.

La recherche des quantiles de la loi du khi-deux à partir de la table de la figure 7.8 est très simple. Supposons que l'on cherche à calculer le quantile $F_X^{-1}(0{,}05)$ d'ordre $\alpha = 5\ \%$ d'une loi $\chi^2(5)$. Par définition :

$$\Pr\left(X \leq F_X^{-1}(\alpha)\right) = \alpha \Longleftrightarrow \Pr\left(X \geq F_X^{-1}(\alpha)\right) = 1 - \alpha \tag{7.115}$$

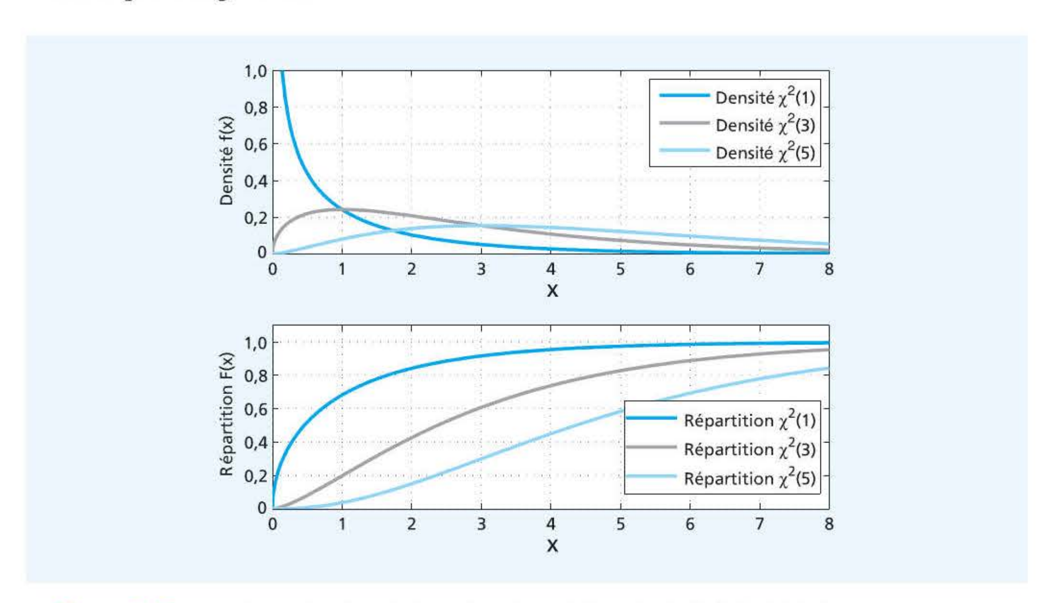
Dans la table, pour une probabilité (probabilité p d'être « supérieur à » indiquée en-têtes de colonnes) égale à $1 - \alpha = 0{,}95$ et une loi du khi-deux à $k = 5$ degrés de liberté

(ligne), on trouve une réalisation égale à 1,145. Par conséquent le quantile d'ordre $\alpha = 5$ % de la loi $\chi^2(5)$ est égal à $F_X^{-1}(0{,}05) = 1{,}145$.

Exemple

Déterminons le quantile d'ordre $\alpha = 90$ % de la loi $\chi^2(5)$. On cherche dans la table la réalisation $F_X^{-1}(0{,}90)$ telle que $\Pr\left(X \geq F_X^{-1}(0{,}90)\right) = 0{,}1$. On trouve $F_X^{-1}(0{,}90) = 9{,}236$.

La figure 7.9 représente les fonctions de densité et de répartition de trois exemples de lois du khi-deux avec des nombres de degrés de liberté respectivement égaux à $k = 1$, $k = 3$ et $k = 5$. On constate que le profil de la densité peut être relativement différent suivant le nombre de degrés de liberté. Pour $k = 1$, la fonction de densité décroit strictement sur $\mathbb{R}^+$, tandis que pour des nombres de degrés de liberté plus élevés apparait une sorte de « bosse » qui tend à se décaler vers la droite au fur et à mesure que k augmente.



▲ **Figure 7.9 Fonctions de densité et de répartition de la loi du khi-deux**

2.4.2 Moments

La fonction génératrice des moments de la loi $\chi^2(k)$ est égale à :

$$M_X(t) = (1-2t)^{-\frac{k}{2}} \quad \forall t < \frac{1}{2} \tag{7.116}$$

De cette fonction génératrice, on peut dériver l'espérance et la variance.

Propriété

Espérance et variance de la loi du khi-deux

Si X suit une loi du khi-deux à k degrés de liberté, alors :

$$\mathbb{E}(X) = k \qquad \mathbb{V}(X) = 2k \tag{7.117}$$

La loi du khi-deux *n'est pas symétrique* par rapport à $\mathbb{E}(X)$, sa skewness est positive ce qui implique que $\Pr(X \geq \mathbb{E}(X)) > \Pr(X \leq \mathbb{E}(X))$. Sa kurtosis est toujours supérieure

à 3 quel que soit le nombre de degrés de liberté. La distribution du khi-deux est une distribution **leptokurtique**.

Propriété

Skewness et kurtosis de la loi du khi-deux

Si $X \sim \chi^2(k)$ alors :

$$\text{Skewness} = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mathbb{E}(X - \mathbb{E}(X))^3)}{\mathbb{V}(X)^{3/2}} = \sqrt{\frac{8}{k}} \tag{7.118}$$

$$\text{Kurtosis} = \frac{\mu_4}{\mu_2^2} = \frac{\mathbb{E}(X - \mathbb{E}(X))^4)}{\mathbb{V}(X)^2} = \frac{12 + 3k}{k} \tag{7.119}$$

2.4.3 Autres propriétés

Propriété

Additivité de la loi du khi-deux

Soient $X_1,\ldots,X_n$ des variables aléatoires réelles indépendantes telles que $X_i \sim \chi^2(k_i)$ pour $i = 1,\ldots,n$. Alors :

$$\sum_{i=1}^{n} X_i \sim \chi^2\left(\sum_{i=1}^{n} k_i\right) \tag{7.120}$$

2.5 Loi de Student

La **loi de Student** (ou distribution *t*), du pseudonyme choisi par William Gosset (1876-1937), est une loi de probabilité continue définie sur l'ensemble des réels $\mathbb{R}$. Cette loi est très utilisée dans la construction d'intervalles de confiance, pour établir la distribution de certaines statistiques de test et notamment du test de Student ou test-*t* (► chapitre 11). La densité d'une loi de Student dépend d'un paramètre appelé nombre de **degrés de liberté**, noté v, avec $v \in \mathbb{N}^*$. Si une variable aléatoire X définie sur $X(\Omega) = \mathbb{R}$ suit une loi de Student à v degrés de liberté, on note alors :

$$X \sim t(v) \tag{7.121}$$

La distribution de Student à v degrés de liberté correspond à la distribution d'un ratio de deux variables indépendantes respectivement distribuées selon une loi normale standard et une loi du khi-deux à v degrés de liberté.

Propriété

Définition d'une loi de Student

Soient Y et Z deux variables aléatoires réelles indépendantes telles que $Y \sim \mathcal{N}(0,1)$ et $Z \sim \chi^2(v)$, alors :

$$\frac{Y}{\sqrt{Z/v}} \sim t(v) \tag{7.122}$$

Puisque la variable normale Y est distribuée sur $\mathbb{R}$ et la variable Z (khi-deux) sur $\mathbb{R}^+$, la loi de Student est définie sur $\mathbb{R}$.

Remarque : Le nombre de dégrés de liberté v est un entier non nul, on ne peut donc pas définir une loi de Student $t(0)$ ou $t(1/2)$ par exemple.

2.5.1 Fonction de densité et fonction de répartition

Définition 7.21

La variable aléatoire réelle X suit une **loi de Student** à $v \in \mathbb{N}^*$ degrés de liberté sur le support $X(\Omega) = \mathbb{R}$ si sa fonction de densité est définie par :

$$f_X(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)}\left(1 + \frac{x^2}{v}\right)^{-\frac{v+1}{2}} \quad \forall x \in \mathbb{R} \qquad (7.123)$$

où $\Gamma(.)$ désigne la *fonction gamma*.

Propriété

Loi de Student

Si la variable aléatoire réelle X suit une loi de Student $t(v)$ alors sa fonction de densité vérifie les propriétés suivantes :

1. $\lim_{x\to+\infty} f_X(x) = \lim_{x\to-\infty} f_X(x) = 0.$
2. $f_X(x) = f_X(-x), \forall x \in \mathbb{R}.$
3. $f_X(x)$ atteint son maximum en $x = 0$.

La première propriété est identique à celle évoquée dans le cas de la loi normale. La seconde propriété signifie que la fonction de densité de la loi normale est symétrique par rapport à son espérance égale à 0. La troisième propriété implique que le mode de la distribution de Student est égal à son espérance.

Comme pour la loi normale et la loi du khi-deux, il n'existe pas de forme analytique pour la fonction de répartition de la loi de Student (on peut utiliser la fonction *LOI.STUDENT* sous Excel). Mais cette fonction de répartition vérifie les propriétés suivantes.

Propriété

Fonction de répartition de la loi de Student

Puisque la densité de la loi de Student est symétrique par rapport à son espérance égale à 0, sa fonction de répartition vérifie :

$$F_X(0) = 0{,}5 \qquad (7.124)$$

$$F_X(-x) = 1 - F_X(x) \quad \forall x \in \mathbb{R} \qquad (7.125)$$

Une table de la fonction de répartition de la loi de Student est fournie en annexe (▸ figure 7.10 pour un extrait). La plupart du temps on utilise les tables de la loi de Student pour déterminer des probabilités du type $\Pr(|X| \geq x)$ (▸ chapitre 11). C'est pourquoi sur la table de la figure 7.10 sont reportées les réalisations x et les probabilités p telles que $\Pr(|X| \geq x) = p$. Les probabilités p figurent en en-têtes de colonnes

et vont de 0,005 à 0,90. En ligne figure le nombre de degrés de liberté v de la loi de Student qui varie de 1 à 20.

r	P= 0,90	P= 0,80	P= 0,70	P= 0,60	P= 0,50	P= 0,40	P= 0,30	P= 0,20	P= 0,10	P= 0,05	P= 0,01	P= 0,005
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	63,657	127,321
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	9,925	14,089
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	5,841	7,453
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	4,604	5,598
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	4,032	4,773
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,707	4,317
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	3,499	4,029
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	3,355	3,833
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	3,250	3,690
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	3,169	3,581
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	3,106	3,497
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	3,055	3,428
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	3,012	3,372
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,977	3,326
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,947	3,286
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,921	3,252
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,898	3,222
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,878	3,197
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,861	3,174
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,845	3,153

▲ **Figure 7.10 Extrait de la table de la loi de Student**

Comment déterminer la valeur de $F_X(x)$ à partir des probabilités $\Pr(|X| \geq x) = p$ reportées dans la table ? De façon générale, on sait que $\Pr(a < X < b) = \Pr(X < b) - \Pr(X < a)$. On en déduit que si $x \geq 0$:

$$\Pr(|X| \geq x) = 1 - \Pr(|X| \leq x) = 1 - \Pr(-x \leq X \leq x)$$

$$= 1 - \Pr(X \leq x) + \Pr(X \leq -x) \quad (7.126)$$

$$= 1 - F_X(x) + F_X(-x) \quad (7.127)$$

Puisque la loi de Student est symétrique, $F_X(-x) = 1 - F_X(x)$. On en déduit que :

$$p = \Pr(|X| \geq x) = 2 - 2F_X(x) \quad (7.128)$$

Par inversion, il vient si $x \geq 0$:

$$F_X(x) = \Pr(X \leq x) = \frac{2 - \Pr(|X| \geq x)}{2} \quad (7.129)$$

Dans le cas $x < 0$, on a :

$$F_X(x) = 1 - F_X(-x) = \frac{\Pr(|X| \geq -x)}{2} \quad (7.130)$$

Exemple

On suppose que $X \sim t(4)$ et que l'on souhaite calculer les probabilités cumulées $\Pr(X \leq 2{,}776)$ et $\Pr(X \leq -0{,}271)$. Nous savons que :

$$F_X(2{,}776) = \Pr(X \leq 2{,}776) = \frac{2 - \Pr(|X| \geq 2{,}776)}{2} \quad (7.131)$$

Sur la ligne $v = 4$ on cherche la réalisation la plus proche de 2,776. Cette réalisation figure dans le tableau et correspond à une probabilité $p = \Pr(|X| \geq 2{,}776)$ égale à 0,05. Par conséquent :

$$F_X(2{,}776) = \frac{2 - 0{,}05}{2} = 0{,}975 \quad (7.132)$$

De la même façon, puisque la réalisation $-0{,}271$ est négative :

$$F_X(-0{,}271) = \Pr(X \leq -0{,}271) = \frac{\Pr(|X| \geq 0{,}271)}{2} \tag{7.133}$$

Sur la ligne $v = 4$ on cherche la réalisation la plus proche de 0,271. On lit dans la table $p = \Pr(|X| \geq 0{,}271) = 0{,}80$, on en déduit que :

$$F_X(-0{,}271) = \frac{0{,}80}{2} = 0{,}40 \tag{7.134}$$

En ce qui concerne la recherche des *quantiles* de la loi de Student à partir de la table de la 7.10, on doit distinguer deux cas :

Premier cas. Si l'on cherche une réalisation c telle que $\Pr(|X| \geq c) = p$, on peut la trouver directement dans la table puisque cette dernière a été conçue pour cela. Par exemple, pour une loi $t(4)$ et une probabilité $p = 10\ \%$ on lit directement $\Pr(|X| \geq 2{,}132) = 0{,}10$.

Deuxième cas. Si l'on cherche un quantile $F_X^{-1}(\alpha)$ d'ordre α, on doit alors utiliser une formule de passage entre ce quantile et la probabilité $p = \Pr\left(|X| \geq F_X^{-1}(\alpha)\right)$. Dans la table, on cherche la réalisation $F_X^{-1}(\alpha)$ telle que :

$$p = \Pr\left(|X| \geq F_X^{-1}(\alpha)\right) = 2 - 2\alpha \quad \text{si } \alpha \geq 0{,}5 \tag{7.135}$$

$$p = \Pr\left(|X| \geq -F_X^{-1}(\alpha)\right) = 2\alpha \quad \text{si } \alpha \leq 0{,}5 \tag{7.136}$$

Exemple

On suppose que $X \sim t(4)$ et que l'on souhaite calculer les quantiles $F_X^{-1}(0{,}05)$ et $F_X^{-1}(0{,}90)$. Pour $\alpha = 5\ \%$, on cherche la réalisation x dans la table de la figure 7.10 correspondant à une probabilité $p = \Pr(|X| \geq x) = 2\alpha = 0{,}1$. Pour la ligne $v = 4$ degrés de liberté et une probabilité $p = 0{,}1$, on lit une réalisation 2,132. Par conséquent, $F_X^{-1}(0{,}05) = -2{,}132$. De la même façon pour $\alpha = 90\ \%$ on cherche la réalisation correspondant à une probabilité $p = 2 - 2\alpha = 0{,}2$. On trouve dans la table $F_X^{-1}(0{,}90) = 1{,}533$.

Propriété

Quantiles de la loi de Student

Puisque la densité de la loi de Student est symétrique par rapport à son espérance égale à 0, sa fonction de répartition inverse vérifie :

$$F_X^{-1}(0{,}5) = 0 \qquad F_X^{-1}(x) \leq 0 \ \text{ si } x \leq 0{,}5 \tag{7.137}$$

$$F_X^{-1}(\alpha) = -F_X^{-1}(1-\alpha) \quad \forall \alpha \in [0{,}1] \tag{7.138}$$

Par exemple, pour une loi $t(4)$ on montre que $F_X^{-1}(0{,}05) = -2{,}132$ et $F_X^{-1}(0{,}95) = 2{,}132$.

2.5.2 Moments

La fonction génératrice des moments de la loi de Student $t(v)$ n'est pas définie. On peut toutefois dériver certains de ses moments.

Propriété

Espérance et variance de la loi de Student

Si X suit une loi de Sudent à $v \in \mathbb{N}^*$ degrés de liberté, alors :

$$\mathbb{E}(X) = 0 \quad \text{si } v \geq 2 \tag{7.139}$$

$$\mathbb{V}(X) = \frac{v}{v-2} \quad \text{si } v \geq 3 \tag{7.140}$$

Nous avions vu dans le chapitre 6 que certains moments de certaines lois de probabilité n'étaient pas définis. C'est le cas de la loi de Student. L'espérance $\mathbb{E}(X)$ d'une loi $t(1)$ n'est pas définie. La variance $\mathbb{V}(X)$ d'une loi $t(1)$ ou $t(2)$ n'existe pas.

Propriété

Skewness et kurtosis de la loi de Student

Si $X \sim t(v)$ alors :

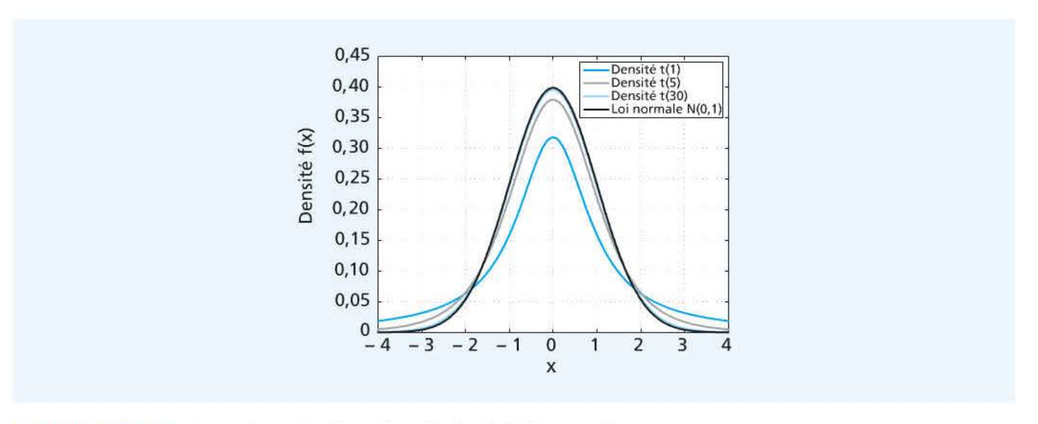
$$\text{Skewness} = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mathbb{E}(X - \mathbb{E}(X))^3)}{\mathbb{V}(X)^{3/2}} = 0 \quad \text{si } v \geq 4 \tag{7.141}$$

$$\text{Kurtosis} = \frac{\mu_4}{\mu_2^2} = \frac{\mathbb{E}(X - \mathbb{E}(X))^4)}{\mathbb{V}(X)^2} = 3 + \frac{6}{v-4} \quad \text{si } v \geq 5 \tag{7.142}$$

La loi de Student est une loi *symétrique* par rapport à son espérance $\mathbb{E}(X)$ (si cette dernière existe). Par ailleurs, une propriété intéressante de la loi de Student est que sa kurtosis (lorsqu'elle existe) dépend du nombre de degrés de liberté. Plus précisément, la kurtosis est une *fonction décroissante* du nombre de degrés de liberté. Plus v diminue, plus la kurtosis augmente et plus la probabilité d'apparition de réalisations extrêmes croît. C'est pourquoi cette loi est beaucoup utilisée pour représenter l'apparition de rendements financiers extrêmes (gains et pertes) sur les marchés financiers. Pour toute valeur de $v \geq 4$ finie, la kurtosis est supérieure à 3 : la distribution de Student est **leptokurtique**. Mais lorsque le nombre de degrés de liberté v tend vers l'infini, la kurtosis tend vers 3 : la distribution est dans ce cas **mesokurtique**.

2.5.3 Autres propriétés

La figure 7.11 représente les fonctions de densité de trois exemples de lois de Student avec des nombres de degrés de liberté respectivement égaux à $v = 1$, $v = 5$ et $v = 30$. Par comparaison, nous reportons également la fonction de densité de la loi normale centrée réduite $\mathcal{N}(0,1)$. On constate que plus le degré de liberté v augmente, plus *la densité de la loi de Student $t(v)$ tend à se rapprocher de celle de la loi normale standard*. On vérifie notamment que la distribution devient de moins en moins leptokurtique au fur et à mesure que v augmente, c'est-à-dire que l'épaisseur de ses queues de distribution tend à diminuer lorsque v augmente.



▲ **Figure 7.11** Fonction de densité de la loi de Student

Propriété

Convergence vers la loi normale

Lorsque le nombre de degrés de liberté tend vers l'infini, la distribution de Student **converge (en distribution)** vers la loi normale centrée réduite.

$$\lim_{\upsilon\to\infty} f_X(x) = \phi(x) \quad \forall x \in \mathbb{R}$$

Nous étudierons les notions de convergence ensuite. Mais nous comprenons d'ores et déjà que ce résultat signifie que lorsque $\upsilon \to \infty$, la loi de probabilité de Student est identique à celle d'une loi normale $\mathcal{N}(0,1)$: sa densité est la même, sa fonction de répartition est la même, ses quantiles sont les mêmes, ses moments sont les mêmes, etc. Par exemple, nous savons que si Y suit une loi normale centrée réduite alors $\mathbb{E}(Y) = 0$, $\mathbb{V}(Y) = 1$, skewness$(Y) = 0$ et kurtosis$(Y) = 3$. Or, si X suit une loi $t(\upsilon)$, il vient :

$$\mathbb{E}(X) = 0 \quad \text{skewness}(X) = 0 \tag{7.143}$$

$$\lim_{\upsilon\to\infty} \mathbb{V}(X) = \lim_{\upsilon\to\infty} \frac{\upsilon}{\upsilon - 2} = 1 \tag{7.144}$$

$$\lim_{\upsilon\to\infty} \text{kurtosis}(Y) = \lim_{\upsilon\to\infty} \left(3 + \frac{6}{\upsilon - 4}\right) = 3 \tag{7.145}$$

On retrouve les mêmes moments que ceux de la loi normale standard.

2.6 Loi de Fisher-Snedecor

La **loi de Fisher-Snedecor** (ou loi de Fisher), du nom des statisticiens britannique Ronald Fisher (1890-1962) et américain George Snedecor (1881-1974), est une loi de probabilité continue définie sur l'ensemble des réels positifs $\mathbb{R}^+$. Cette loi est notamment utilisée pour caractériser la distribution de certaines statistiques de test sous l'hypothèse nulle et en particulier celle du test de Fisher ou F-test (▶ chapitre 11). La densité d'une loi de Fisher dépend de deux paramètres n et m qui correspondent à des nombres de **degrés de liberté**, avec $(n,m) \in \mathbb{N}^* \times \mathbb{N}^*$. Si une variable aléatoire X définie sur $X(\Omega) = \mathbb{R}^+$ suit une loi de Fisher à n et m degrés de liberté, on note :

$$X \sim \mathcal{F}(n,m) \tag{7.146}$$

Propriété

Distribution d'une loi de Fisher

Soient Y et Z deux variables aléatoires réelles indépendantes telles que $Y \sim \chi^2(n)$ et $Z \sim \chi^2(m)$, alors :

$$\frac{Y/n}{Z/m} \sim \mathcal{F}(n,m) \tag{7.147}$$

Remarque : Une loi de Fisher étant définie comme la loi du ratio de deux variables distribuées selon une loi du khi-deux, elle est définie sur $\mathbb{R}^+$: la réalisation d'une variable de Fisher n'est jamais négative.

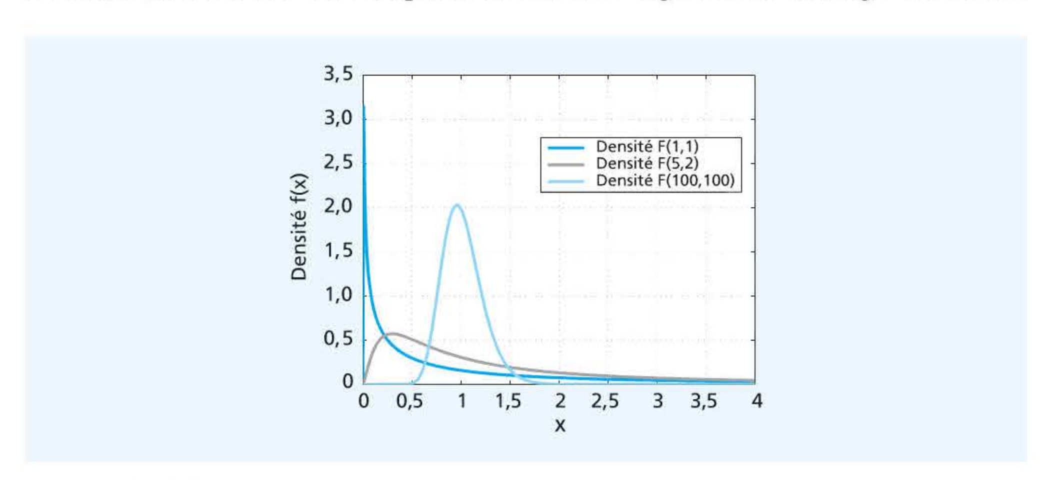
Définition 7.22

La variable aléatoire réelle X suit une **loi de Fisher** à $n \in \mathbb{N}^*$ et $m \in \mathbb{N}^*$ degrés de liberté sur le support $X(\Omega) = \mathbb{R}^+$ si sa fonction de densité est définie par :

$$f_X(x) = \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)}{\Gamma\left(\frac{n}{2}+\frac{m}{2}\right)} n^{\frac{n}{2}} m^{\frac{m}{2}} \frac{x^{\frac{n}{2}-1}}{(m+nx)^{\frac{n+m}{2}}} \quad \forall x \in \mathbb{R}^+ \tag{7.148}$$

où $\Gamma(.)$ désigne la *fonction gamma*.

La figure 7.12 représente les fonctions de densité de trois exemples de lois de Fisher avec des nombres de degrés de liberté respectivement égaux à (1,1), (5,2), (100,100). On constate que, suivant les valeurs des nombres de degrés de liberté n et m, on peut obtenir des profils très variés de la fonction de densité : strictement décroissante dans le cas (1,1), avec une bosse lorsque les nombres de degrés de liberté augmentent, etc.



▲ **Figure 7.12 Fonction de densité de la loi de Fisher**

Étant donnée la définition de la fonction de densité, on vérifie que si X a une distribution de Fisher, alors $1/X$ a aussi une distribution de Fisher.

Propriété

Loi de Fisher

Soit X une variable aléatoire réelle telle que $X \sim \mathcal{F}(n,m)$ alors $X^{-1} \sim \mathcal{F}(m,n)$.

Par conséquent, si l'on note $F_X(x; n,m)$ la fonction de répartition de la loi $\mathcal{F}(n,m)$, on a une relation du type :

$$F_X(x; n,m) = 1 - F_X\left(\frac{1}{x}; m,n\right) \quad \forall x \in \mathbb{R}^+ \tag{7.149}$$

De la même façon si l'on note $F_X^{-1}(\alpha; n,m)$ le quantile d'ordre α de la loi $\mathcal{F}(n,m)$, il vient :

$$F_X^{-1}(\alpha; n,m) = \frac{1}{F_X^{-1}(1-\alpha; m,n)} \quad \forall \alpha \in [0,1] \tag{7.150}$$

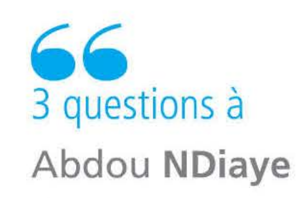
La fonction de répartition de la loi de Fisher n'a pas d'expression analytique. Une table statistique fournie en annexe de cet ouvrage (▶ figure 7.13 pour un extrait) permet de retrouver les probabilités cumulées $F_X(x) = \Pr(X \leq x)$ d'une loi de Fisher à v_1 (en ligne) et v_2 (en colonne) degrés de liberté. Sur cette table sont reportées les réalisations x telles que $\Pr(X \geq x) = p$ pour deux niveaux de probabilité $p = 0{,}05$ et $p = 0{,}01$. Par exemple, on lit que pour une loi $\mathcal{F}(3,4)$ on a :

$$F_X(6{,}591) = \Pr(X \leq 6{,}591) = 1 - \Pr(X \geq 6{,}591) = 1 - 0{,}05 = 0{,}95 \tag{7.151}$$

	$v_1 = 1$		$v_1 = 2$		$v_1 = 3$		$v_1 = 4$		$v_1 = 5$	
v_2	P= 0,05	P= 0,01	P= 0,05	P= 0,01	P= 0,05	P= 0,01	P= 0,05	P= 0,01	P= 0,05	P= 0,01
1	161,448	4052,181	199,500	4999,500	215,707	5403,352	224,583	5624,583	230,162	5763,650
2	18,513	98,503	19,000	99,000	19,164	99,166	19,247	99,249	19,296	99,299
3	10,128	34,116	9,552	30,817	9,277	29,457	9,117	28,710	9,013	28,237
4	7,709	21,198	6,944	18,000	6,591	16,694	6,388	15,977	6,256	15,522
5	6,608	16,258	5,786	13,274	5,409	12,060	5,192	11,392	5,050	10,967
6	5,987	13,745	5,143	10,925	4,757	9,780	4,534	9,148	4,387	8,746
7	5,591	12,246	4,737	9,547	4,347	8,451	4,120	7,847	3,972	7,460
8	5,318	11,259	4,459	8,649	4,066	7,591	3,838	7,006	3,687	6,632
9	5,117	10,561	4,256	8,022	3,863	6,992	3,633	6,422	3,482	6,057
10	4,965	10,044	4,103	7,559	3,708	6,552	3,478	5,994	3,326	5,636
11	4,844	9,646	3,982	7,206	3,587	6,217	3,357	5,668	3,204	5,316
12	4,747	9,330	3,885	6,927	3,490	5,953	3,259	5,412	3,106	5,064
13	4,667	9,074	3,806	6,701	3,411	5,739	3,179	5,205	3,025	4,862
14	4,600	8,862	3,739	6,515	3,344	5,564	3,112	5,035	2,958	4,695
15	4,543	8,683	3,682	6,359	3,287	5,417	3,056	4,893	2,901	4,556
16	4,494	8,531	3,634	6,226	3,239	5,292	3,007	4,773	2,852	4,437
17	4,451	8,400	3,592	6,112	3,197	5,185	2,965	4,669	2,810	4,336
18	4,414	8,285	3,555	6,013	3,160	5,092	2,928	4,579	2,773	4,248
19	4,381	8,185	3,522	5,926	3,127	5,010	2,895	4,500	2,740	4,171
20	4,351	8,096	3,493	5,849	3,098	4,938	2,866	4,431	2,711	4,103

▲ **Figure 7.13** **Extrait de la table de la loi de Fisher**

Sur cette table le nombre de degrés de libérté v_1 n'est reporté que pour certaines valeurs (1, 2, 3, 4, 5, 6, 8, 10, 12, 24, 48) tandis que v_2 est reporté pour toutes les valeurs allant de 1 à 30, puis pour les valeurs 40 et 80. Si l'on souhaite calculer la probabilité d'une loi de Fisher dont les nombres de degrés de liberté v_1 et/ou v_2 ne sont pas reportés dans cette table, on peut utiliser la relation de l'équation (7.149) pour obtenir le résultat.



3 questions à

Abdou **NDiaye**

Risk Management Analyst,
Volkswagen Bank



Quel est votre parcours professionnel et votre mission actuelle chez Volkswagen Bank ?

À l'issue de mes études, j'ai été embauché en 2009 à CGI-Finance, filiale de la Société Générale. Actuellement chez Volkswagen Bank depuis mars 2012, j'ai intégré le département Risk Management comme analyste senior. En tant que statisticien, mes missions sont très variées, mais portent sur deux axes majeurs : la gestion des risques (retail et corporate) et le marketing financier. Concernant la gestion des risques, je contribue à la modélisation, au backtesting et à l'estimation de modèles de notation interne (scores d'acceptation, scores de comportement, modèles LGD, scores recouvrement). Je travaille en outre sur la mise en œuvre de stress testing et sur la préparation du Business Financial Review (semestriel). Dans le domaine du marketing, mon travail consiste à mettre en œuvre des modèles de suivi pour mesurer le risque de remboursement anticipé, des méthodes de ciblage de clientèle et des études de retour de campagne marketing pour la fidélisation de nos clients.

Dans le cadre de votre activité, quelles lois de probabilité usuelles utilisez-vous et dans quel cadre ?

On utilise les lois de probabilité usuelles à de nombreuses occasions. Ces distributions sont notamment utilisées pour caractériser la loi de différentes statistiques de tests. Par exemple, nous utilisons des tests sur les liaisons entre la variable « cible » et les variables explicatives afin d'éliminer les variables sans aucune influence sur la variable « cible ». Ces tests reposent sur la loi du khi-deux (test du khi-deux par exemple). Mais de façon générale, nous utilisons régulièrement différentes lois usuelles comme la loi normale (pour les modèles probit), la loi de Student, la loi logistique (pour les modèles logit), la loi exponentielle (pour les durées), etc.

Sous quels types de logiciels la mise en œuvre de ces lois se fait-elle ?

Uniquement sur des logiciels statistiques : à la Société Générale sous SAS et chez Volkswagen Bank sous SPSS. ■

Les points **clés**

→ La loi uniforme discrète est caractérisée par une propriété d'équiprobabilité.

→ La loi de Bernoulli est la loi de probabilité de variables aléatoires dichotomiques ou binaires.

→ La loi binomiale correspond à la loi de probabilité du nombre de succès obtenus lors de n expériences indépendantes de Bernoulli de même probabilité de succès.

→ La loi géométrique correspond à la loi de probabilité du rang du premier succès obtenu lors de n expériences indépendantes de Bernoulli de même probabilité de succès.

→ La loi de Poisson est une loi de probabilité adaptée aux variables de comptage.

→ La loi exponentielle et la loi géométrique sont des lois de probabilité sans mémoire.

→ La loi normale est une loi de probabilité continue et symétrique définie sur $\mathbb{R}$.

→ La loi du khi-deux correspond à la loi de probabilité d'une somme de carrés de variables normales centrées, réduites et indépendantes.

→ La loi de Student correspond à la loi de probabilité d'un ratio de variables indépendantes distribuées selon une loi normale standard et une loi du khi-deux.

→ La loi de Fisher correspond à la loi de probabilité d'un ratio de variables indépendantes distribuées selon des lois du khi-deux.

ÉVALUATION

► Corrigés sur www.dunod.com

QCM

Pour chacune des questions suivantes, indiquer si les affirmations sont vraies ou fausses (il peut y avoir plusieurs réponses vraies pour chaque question).

1 Lois usuelles discrètes

a. La loi binomiale correspond au résultat du $n^{ème}$ tirage d'une expérience de Bernoulli.

b. Une somme de variables de Bernoulli indépendantes est distribuée selon une loi binomiale.

c. Une somme de variables indépendantes distribuées selon des lois binomiales de même probabilité de succès est distribuée selon une loi binomiale.

d. Une loi de Poisson permet de modéliser des variables de comptage.

e. Une loi géométrique est une loi sans mémoire.

2 Lois usuelles continues

a. La fonction de densité d'une loi uniforme continue est symétrique par rapport à son espérance.

b. Une loi uniforme est une loi de probabilité sans mémoire.

c. La fonction de densité d'une loi exponentielle est symétrique par rapport à son espérance.

d. Si une variable suit une loi normale $\mathcal{N}\left(\mu,\sigma^2\right)$, son carré suit une loi du khi-deux à 1 degré de liberté.

e. La loi de Fisher-Snedecor correspond à la loi d'un ratio de variables indépendantes distribuées selon des lois du khi-deux.

3 Loi normale

a. Si X suit une loi normale alors la variable $-3\times X$ suit une loi normale.

b. Si $\Phi(1{,}96) = 0{,}975$, alors $\Phi^{-1}(0{,}025) = 1{,}96$.

c. Si $\alpha_1 > \alpha_2$ alors $\Phi^{-1}(\alpha_1) > \Phi^{-1}(\alpha_2)$.

d. Si $\Phi(-1{,}96) = 0{,}025$ alors $\Phi(1{,}96) = 0{,}975$.

e. La densité de la loi normale est symétrique autour de son espérance.

4 Loi de Student

a. La fonction de densité de la loi de Student est toujours leptokurtique.

b. La fonction de densité de la loi de Student est unimodale et son mode est égal à son espérance.

c. Si $F_X(x)$ est la fonction de répartition d'une loi $t(v)$, alors $F_X(x) = -F_X(1-x)$, $\forall x \in \mathbb{R}$.

d. Si $F_X(x)$ est la fonction de répartition d'une loi $t(v)$, alors $F_X^{-1}(\alpha) = -F_X^{-1}(1-\alpha)$, $\forall \alpha \in [0{,}1]$.

e. Si $F_X(x)$ est la fonction de répartition d'une loi $t(4)$, alors $F_X^{-1}(0{,}05) = 1{,}95$.

Exercices

5 Lois usuelles discrètes

L'objectif de cet exercice est de vous familiariser avec le calcul des probabilités, des probabilités cumulées et des fractiles des principales lois discrètes usuelles à partir des tables statistiques fournies en annexe.

1. Soit X une variable aléatoire distribuée selon une loi uniforme discrète sur $X(\Omega) = \{-2,-1,0,1,2,3\}$. Calculez $\Pr(X = 4)$, $\Pr(X = 1)$ et $\Pr(X \leq 2)$.

2. Soit X une variable aléatoire discrète distribuée selon une loi binomiale $\mathcal{B}(10\ ;\ 0{,}4)$. Calculez $\Pr(X = 2)$, $\Pr(X = 2{,}57)$ et $\Pr(2{,}57 < X \leq 4)$.

3. Soit X une variable aléatoire discrète distribuée selon une loi de Poisson $\mathcal{P}(4)$. Calculez $\Pr(X = 2{,}57)$, $\Pr(X = 5)$ et $\Pr(2{,}57 < X \leq 4)$.

6 Loi normale

L'objectif de cet exercice est de vous familiariser avec le calcul des probabilités cumulées et des fractiles de la loi normale à partir des tables statistiques fournies en annexe. Soit X une variable aléatoire continue distribuée selon une loi normale $\mathcal{N}(2{,}2)$.

1. Calculez $\Pr(X = 2{,}57)$ et $\Pr(|X| \geq 1)$.

2. Calculez les quantiles d'ordres $\alpha = 0$ et $\alpha = 0{,}95$.

3. Calculez le quantile d'ordre $\alpha = 0{,}01$.

7 Loi binomiale (Université Lyon 1, 2008-09)

On extrait $n \in \mathbb{N}$ fois avec remise une boule dans une urne composée de 2 boules vertes et 6 boules blanches. Soit X_n la variable aléatoire associée au nombre de boules vertes obtenues lors des n tirages.
On pose $F_n = X_n/n$.

1. Donner la loi de X_n. En déduire l'espérance et la variance de X_n puis de F_n.

2. On suppose dans cette question que $n = 10\,000$. A l'aide de l'exercice précédent, donner une borne inférieure pour la probabilité de l'événement

$$F_n \in]0{,}22; 0{,}26[\qquad (7.152)$$

3. Déterminer le nombre minimal n de tirages nécessaires pour que la probabilité de l'événement $F_n \in]0{,}22; 0{,}26[$ soit au moins égale à 0,99.

Sujets d'examen

8 Loi normale (Université Paris Assas)

Soit X une variable aléatoire dont la loi de probabilité est la loi normale de moyenne 12 et d'écart-type 4.

1. Calculer la probabilité de réalisation de chacun des événements suivants :

$$(X = 2) \quad (X < 16) \quad (X > 20) \quad (X < 0) \qquad (7.153)$$

2. Déterminer le nombre e tel que la probabilité de réalisation de l'événement $(|X - 12| > e)$ soit égale à 0,01.

3. Soit Y la variable définie par la relation suivante :

$$Y = aX + b \qquad (7.154)$$

où a et b désignent deux paramètres réels. Calculer en fonction de a et de b, l'espérance mathématique et la variance de Y. Déterminez a et b sachant que la probabilité de réalisation de l'événement $(Y < 24)$ est égale à 0,2266 et que celle de l'événement $(Y > 42)$ est égale à 0,0668.

9 Loi binomiale et loi de Poisson (Université Paris Assas)

On considère les nombres x_0 et x_1 définis comme suit :

– x_0 est la plus grande valeur entière de x telle que :

$$\Pr(X \leq x) \leq 0{,}05 \qquad (7.155)$$

– x_1 est la plus petite valeur entière de x telle que :

$$\Pr(X \geq x) \leq 0{,}05 \qquad (7.156)$$

1. Déterminer les valeurs x_0 et x_1 si la variable X suit une loi binomiale $\mathcal{B}(40; 0{,}08)$.

2. Déterminer les valeurs x_0 et x_1 si la variable X suit une loi de Poisson $\mathcal{P}(12)$.

10 Loi de Fisher et loi du khi-deux (Université Paris Assas)

On considère les nombres x_0 et x_1 définis comme suit :

$$\Pr(X < x_0) = 0{,}05 \qquad (7.157)$$

$$\Pr(X > x_1) = 0{,}05 \qquad (7.158)$$

1. Déterminer les valeurs x_0 et x_1 si la variable X suit une loi de Fisher $\mathcal{F}(4,20)$.

2. Déterminer les valeurs x_0 et x_1 si la variable X suit une loi du khi-deux $\chi^2(8)$.

Chapitre 8

L'assurance est une activité fondamentalement basée sur la loi des grands nombres. Considérons le cas d'une assurance automobile destinée à couvrir le risque de dégradation du véhicule (accident, vol, bris de glace, etc.). Pour un individu donné, la perte financière associée à un tel risque peut être représentée par une variable aléatoire. Supposons que l'espérance de cette variable soit égale à 170 €. Si chaque individu accepte de verser à la compagnie d'assurance une somme légèrement supérieure à 170 €, par exemple 175 €, le montant total des primes collectés sera de 175n € où n désigne le nombre d'assurés. Supposons que les pertes des assurés sont indépendantes. Sous ces hypothèses, la loi des grands nombres permet de montrer que si n est suffisamment grand, le montant total des pertes sera inférieur au total des primes collectées avec une probabilité de 1, ce qui garantit la pérennité financière de la compagnie d'assurance.

LES GRANDS AUTEURS



Jarl Waldemar Lindeberg (1876-1932)



Jarl Waldemar Lindeberg est un mathématicien finnois qui fut à l'origine de l'une des versions du théorème central limite, la version dite de Lindeberg-Levy que nous découvrirons dans ce chapitre. Cette version du théorème est parue dans un article en 1920, quelques temps après la publication d'une autre version proposée par le mathématicien russe Alexandre Lyapounov (1857-1918). Lindeberg dit ne pas avoir eu connaissance des travaux de Lyapounov, ce qui peut paraître curieux à l'ère d'Internet, mais ce qui se comprend tout à fait dans le contexte du début du XX[e] siècle. D'ailleurs, comme nous le verrons dans ce chapitre, leurs approches sont très différentes.

Le mathématicien suédois Harald Cramer rapporte une anecdote plaisante au sujet de Lindeberg : quand on venait à lui reprocher de ne pas être suffisamment actif dans son travail de recherche, il répondait qu'il s'occupait principalement de sa ferme. Mais quand on lui reprochait un certain laisser-aller dans son exploitation agricole, il répondait qu'il était avant tout professeur. Quoiqu'il en soit, la moyenne de ses compétences agricoles et scientifiques est passée à la postérité. ■

Propriétés asymptotiques

Plan

Pré-requis

→ **Connaître** la notion de variable aléatoire (▸ chapitre 6).

Objectifs

→ **Étudier** le comportement asymptotique d'une suite de variables aléatoires.

→ **Comprendre** le concept de convergence en probabilité.

→ **Comprendre** le concept de convergence presque sûre.

→ **Comprendre** le concept de convergence en loi.

→ **Savoir appliquer** la loi faible des grands nombres.

→ **Savoir appliquer** le théorème central limite.

n considère n variables aléatoires (discrètes ou continues) $Y_1, Y_2, \ldots, Y_n$. À partir de celles-ci, on construit une nouvelle variable, notée X_n, telle que X_n soit une fonction $f(.)$ de $Y_1, Y_2, \ldots, Y_n$:

$$X_n = f(Y_1, \ldots, Y_n) \tag{8.1}$$

Le but de chapitre est d'étudier le comportement de la **suite de variables aléatoires** X_n lorsque la dimension n tend vers l'infini. Est-ce que X_n est toujours définie comme une variable aléatoire lorsque n tend vers l'infini ? Ou, au contraire, se comporte-t-elle comme une variable dégénérée (quantité certaine) ? Quelle est la **loi asymptotique** de cette variable ?

Pour répondre à ces questions, nous allons introduire différents concepts de **convergence**. La notion de convergence constitue la base de la statistique mathématique et de la théorie des tests. Dans ce cadre, nous présenterons deux résultats fondamentaux : la **loi des grands nombres** et le **théorème central limite**. Ces deux théorèmes s'intéressent au comportement asymptotique d'une fonction particulière des variables $Y_1, \ldots, Y_n$, à savoir la **moyenne empirique** :

$$f(Y_1, \ldots, Y_n) = \frac{1}{n} \sum_{i=1}^{n} Y_n \tag{8.2}$$

Pourquoi s'intéresser tout particulièrement au comportement asymptotique de la moyenne empirique ? Dans la pratique, la variable étudiée correspond généralement à un **estimateur** (► chapitre 9) et la dimension n à la taille de l'échantillon. On souhaite alors étudier les **propriétés asymptotiques** de cet estimateur en faisant tendre la taille de l'échantillon vers l'infini. Or, la plupart des estimateurs peuvent s'écrire comme des fonctions de la moyenne empirique des variables de l'échantillon. Dès lors, la connaissance du comportement asymptotique de la moyenne empirique permet d'en déduire le comportement asymptotique de la plupart des estimateurs. Ces deux théorèmes constituent ainsi le fondement de toute la théorie de l'estimation.

1 Notions de convergence

L'objectif de cette section est d'analyser le comportement d'une **suite (ou séquence) de variables aléatoires**, indicées par $n \in \mathbb{N}$. Comme pour les suites de nombres, une suite de variables aléatoires est une famille de variables aléatoires indexée par un entier strictement positif. Une suite de variables aléatoires est généralement notée sous la forme $(X_n)_{n \in \mathbb{N}}$ ou (X_n). Souvent en statistique, cette suite est définie comme une fonction $X_n = f(Z_1, \ldots, Z_n)$ d'autres variables aléatoires $Z_1, \ldots, Z_n$. C'est typiquement le cas des estimateurs (► chapitre 9). Par exemple, (X_n) peut correspondre à une moyenne empirique, une variance empirique, etc.

Définition 8.1

La **théorie asymptotique** consiste en l'étude des propriétés d'une suite de variables aléatoires (X_n) lorsque la dimension n tend vers l'infini.

La théorie asymptotique est donc basée sur l'idée de limite : on cherche à étudier la « *limite* » de la variable aléatoire X_n lorsque $n \to \infty$. Plus spécifiquement, elle repose sur quatre **notions de convergence** :

1. La convergence **presque sûre**.
2. La convergence en **probabilité**.
3. La convergence en **moyenne quadratique**.
4. La convergence en **loi**.

Remarque : Les notions de convergence peuvent s'appliquer à des suites de variables aléatoires *discrètes* ou *continues*. Afin de simplifier la présentation, nous ne considérerons que le cas de suites de variables aléatoires réelles (continues).

1.1 Convergence presque sûre

La **convergence presque sûre** implique que lorsque n tend vers l'infini, la suite de variables aléatoires (X_n) tend vers une constante déterministe de façon certaine.

Définition 8.2

La suite de variables aléatoires (X_n) **converge presque sûrement** (ou de façon forte) vers une constante c, si :

$$\Pr\left(\lim_{n\to\infty} X_n = c\right) = 1 \quad (8.3)$$

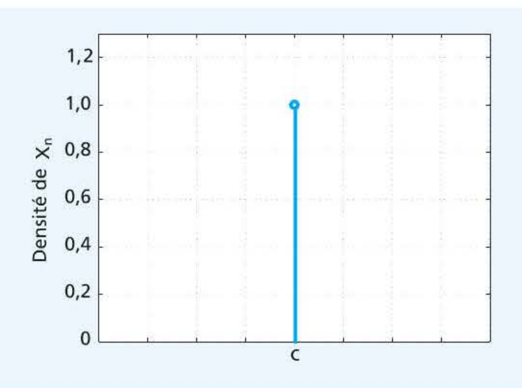
où Pr désigne la probabilité. On note la convergence presque sûre sous la forme suivante :

$$X_n \overset{a.s.}{\to} c \quad (8.4)$$

Le symbole « a.s. » renvoie à la traduction anglo-saxonne du terme presque sûre (*almost sure*). Ce résultat signifie que lorsque n tend vers l'infini, les réalisations de la variable X_n sont toutes égales à la constante c. Par exemple le résultat $X_n \overset{a.s.}{\to} 2$ signifie que pour une dimension n suffisamment grande, si l'on effectue des tirages de la variable X_n, on obtiendra 2, 2,...,2. En d'autres termes, la suite (X_n) tend vers une variable **aléatoire dégénérée**, c'est-à-dire une quantité déterministe (non aléatoire). Comme le montre la figure 8.1, la fonction de densité de la variable X_n, lorsque n tend vers l'infini, est une masse ponctuelle. La probabilité d'obtenir toute autre valeur que $c = 2$ est alors nulle.

1.2 Convergence en probabilité

L'idée de la **convergence en probabilité** est assez similaire à celle de la convergence presque sûre. Lorsque la dimension n tend vers l'infini, la suite (X_n) tend vers une constante déterministe c. La différence est que cette convergence n'est pas stricte : la variable X_n est *presque* dégénérée, mais elle reste toutefois une variable aléatoire même si sa densité est extrêmement concentrée autour de la valeur c.



▲ **Figure 8.1** **Illustration de la notion de convergence presque sûre**

Définition 8.3

La suite de variables aléatoires (X_n) **converge en probabilité** (ou converge au sens faible) vers une constante c, si pour toute valeur $\varepsilon > 0$,

$$\lim_{n\to\infty} \Pr(|X_n - c| > \varepsilon) = 0 \qquad (8.5)$$

Cette convergence en probabilité est notée :

$$X_n \xrightarrow{p} c \quad \text{ou} \quad \text{plim}\, X_n = c \qquad (8.6)$$

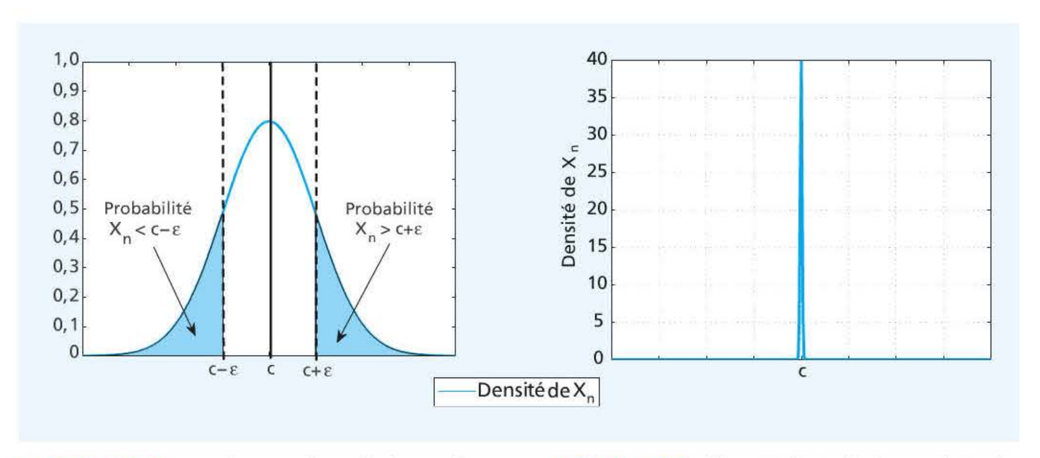
Cette définition signifie que lorsque la dimension n est suffisamment grande, la probabilité d'obtenir une réalisation de X_n dont l'écart à la valeur c (en valeur absolue) soit plus grand qu'une valeur arbitraire ε (aussi petite que l'on souhaite), tend vers 0.

Représentons graphiquement la probabilité $\Pr(|X_n - c| > \varepsilon)$. Cette probabilité correspond à la somme des probabilités associées à deux événements disjoints : soit la variable X_n est supérieure à $c + \varepsilon$, soit la variable X_n est inférieure à $c - \varepsilon$:

$$\Pr(|X_n - c| > \varepsilon) = \Pr(X_n > c + \varepsilon) + \Pr(X_n < c - \varepsilon) \qquad (8.7)$$

Sur la figure 8.2, la probabilité $\Pr(|X_n - c| > \varepsilon)$ est représentée par la somme des aires colorées, situées sous la fonction de densité de X_n, à droite et à gauche des valeurs $c + \varepsilon$ et $c - \varepsilon$. Si $X_n \xrightarrow{p} c$, cela implique que lorsque n tend vers l'infini, la probabilité $\Pr(|X_n - c| > \varepsilon)$ tend vers 0. L'aire colorée de la figure 8.2 devient ainsi très petite. Ceci est vrai quelle que soit la valeur de la constante ε. Pour une quantité ε elle même très petite, la densité de X_n est donc nécessairement **extrêmement concentrée** autour de la valeur c, comme le montre la figure 8.3.

Ainsi le résultat $X_n \xrightarrow{p} 2$ implique que, si l'on effectue des tirages dans X_n pour n très grand, on obtiendra des réalisations du type 2,001, 1,999, 2, 2,0002, etc. La différence entre la notion de convergence presque sûre et la convergence en probabilité est que dans le premier cas, la variable X_n n'est plus une variable aléatoire : sa densité est une masse ponctuelle et ses réalisations sont toujours égales à c. Dans le cas de la convergence en probabilité, X_n n'est presque plus une variable aléatoire : sa densité est extrêmement concentrée autour de c (sa variance est extrêmement faible) et les réalisations de X_n sont très proches de c, mais pas nécessairement égales à c. Par conséquent, la convergence presque sûre *implique* la convergence en probabilité.



▲ **Figure 8.2** **Représentation de la probabilité** $\Pr(|X_n - c| > \varepsilon)$

▲ **Figure 8.3** **Illustration de la notion de convergence en probabilité**

Remarque : La convergence presque sûre implique la convergence en probabilité. Si $X_n \overset{a.s.}{\rightarrow} c$ alors $X_n \overset{p}{\rightarrow} c$, la réciproque n'est pas nécessairement vraie.

FOCUS

Notations

De façon générale, la convergence en probabilité signifie qu'une séquence de variables aléatoires ((X_n) dans notre cas) tend vers une constante lorsque la dimension n est grande :

$$\underbrace{X_n}_{\text{variable aléatoire}} \overset{p}{\rightarrow} \underbrace{c}_{\text{constante}} \tag{8.8}$$

Toutefois, on trouve parfois la notation suivante :

$$\underbrace{X_n}_{\text{variable aléatoire}} \overset{p}{\rightarrow} \underbrace{Z}_{\text{variable aléatoire}} \tag{8.9}$$

où Z est une variable aléatoire non indicée par n. Cela signifie que la différence entre les deux variables X_n et Z tend vers 0 lorsque n tend vers l'infini :

$$\underbrace{X_n - Z}_{\text{variable aléatoire}} \overset{p}{\rightarrow} \underbrace{0}_{\text{constante}} \tag{8.10}$$

Dit autrement, les deux variables ont la même distribution (► section 1.4) lorsque la dimension n est suffisamment grande. Mais attention cette notation peut induire en erreur car, de façon générale, la convergence en probabilité implique une variable aléatoire (à gauche du signe $\overset{p}{\rightarrow}$) et une constante (à droite de ce signe).

Sous quelles conditions, une suite de variables aléatoires X_n converge-t-elle en probabilité ? Une condition nécessaire et suffisante à la convergence en probabilité est la suivante.

Propriété

Convergence en probabilité

Soit une séquence de variables aléatoires (X_n) telle que :

$$\lim_{n\to\infty} \mathbb{E}(X_n) = c \tag{8.11}$$

$$\lim_{n\to\infty} \mathbb{V}(X_n) = 0 \tag{8.12}$$

où $c \in \mathbb{R}$, alors, la suite (X_n) converge en probabilité vers c lorsque $n \to \infty$:

$$X_n \overset{p}{\to} c \tag{8.13}$$

Rappelons que les expressions $\mathbb{E}(X)$ et $\mathbb{V}(X)$ désignent respectivement l'espérance et la variance (▸ chapitre 6) d'une variable aléatoire X.

Exemple

On considère n variables aléatoires $Y_1,\ldots,Y_n$ indépendantes et identiquement distribuées (i.i.d.) telles que $Y_i \sim \mathcal{N}\left(m,\sigma^2\right)$, $\forall i = 1,\ldots,n$. Montrons que la variance empirique corrigée définie par :

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \overline{Y}_n\right)^2 \tag{8.14}$$

où $\overline{Y}_n = (1/n)\sum_{i=1}^{n} Y_i$, converge en probabilité vers la variance σ^2. Pour cela, calculons les deux premiers moments de S_n^2. On admet le résultat suivant :

$$\frac{(n-1)}{\sigma^2}S_n^2 \sim \chi^2(n-1) \quad \forall n \in \mathbb{N} \tag{8.15}$$

Étant données les propriétés de la loi du khi-deux (▸ chapitre 7), on sait que si $X \sim \chi^2(v)$ alors $\mathbb{E}(X) = v$ et $\mathbb{V}(X) = 2v$. Dès lors, on obtient :

$$\mathbb{E}\left(\frac{(n-1)}{\sigma^2}S_n^2\right) = n-1 \tag{8.16}$$

$$\mathbb{V}\left(\frac{(n-1)}{\sigma^2}S_n^2\right) = 2(n-1) \tag{8.17}$$

On en déduit que :

$$\mathbb{E}\left(\frac{(n-1)}{\sigma^2}S_n^2\right) = \frac{(n-1)}{\sigma^2}\mathbb{E}\left(S_n^2\right) = n-1 \Longleftrightarrow \mathbb{E}\left(S_n^2\right) = \sigma^2 \tag{8.18}$$

$$\mathbb{V}\left(\frac{(n-1)}{\sigma^2}S_n^2\right) = \frac{(n-1)^2}{\sigma^4}\mathbb{V}\left(S_n^2\right) = 2(n-1) \Longleftrightarrow \mathbb{V}\left(S_n^2\right) = \frac{2\sigma^4}{n-1} \tag{8.19}$$

Par conséquent :

$$\mathbb{E}\left(S_n^2\right) = \sigma^2 \tag{8.20}$$

$$\lim_{n\to\infty}\mathbb{V}\left(S_n^2\right) = \lim_{n\to\infty}\left(\frac{2\sigma^4}{n-1}\right) = 0 \tag{8.21}$$

On en conclut que la variance empirique corrigée converge vers la variance σ^2.

$$S_n^2 \overset{p}{\to} \sigma^2 \tag{8.22}$$

Une des principales applications de la notion de convergence en probabilité est la **loi faible des grands nombres**. La loi faible des grands nombres, telle qu'énoncée par Khintchine (1878-1959), stipule que la moyenne empirique de variables aléatoires indépendantes et identiquement distribuées (i.i.d.)[1] converge en probabilité vers l'espérance de ces variables.

1 Il existe d'autres versions de la loi faible des grand nombres. Par exemple, il est possible de considérer des variables qui ne sont pas identiquement distribuées, mais seulement indépendantes. Ainsi, la loi faible des grands nombres s'applique même si les espérances ou les variances sont hétérogènes (loi faible des grands nombres de Tchebychev).

Théorème 8.1

Loi faible des grands nombres ou théorème de Khintchine

Soient des variables aléatoires $Y_1,\ldots,Y_n$ indépendantes telles que $\mathbb{E}(Y_i) = m$ et $\mathbb{V}(Y_i) = \sigma^2$, $\forall i = 1,\ldots,n$. La moyenne empirique de ces variables converge en probabilité vers l'espérance m :

$$\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{p} \mathbb{E}(Y_i) = m \tag{8.23}$$

Démonstration

Pour démontrer le résultat de la loi faible des grands nombres, déterminons les deux premiers moments de la moyenne empirique $\overline{Y}_n$. L'espérance étant un opérateur linéaire, nous avons :

$$\mathbb{E}\left(\overline{Y}_n\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Y_i) \tag{8.24}$$

Puisque toutes les variables Y_i ont la même espérance $\mathbb{E}(Y_i) = m$, il vient :

$$\mathbb{E}\left(\overline{Y}_n\right) = \frac{1}{n}\sum_{i=1}^{n} m = \frac{n \times m}{n} = m \tag{8.25}$$

La variance de $\overline{Y}_n$ peut s'écrire sous la forme suivante :

$$\mathbb{V}\left(\overline{Y}_n\right) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\mathbb{V}\left(\sum_{i=1}^{n} Y_i\right) \tag{8.26}$$

En développant, nous faisons apparaître les termes de variances et de covariances des variables $Y_1,\ldots,Y_n$:

$$\mathbb{V}\left(\overline{Y}_n\right) = \frac{1}{n^2}\left(\mathbb{V}(Y_1) + \ldots + \mathbb{V}(Y_n) + 2\mathbb{C}ov(Y_1,Y_2) + \ldots + 2\mathbb{C}ov(Y_{n-1},Y_n)\right) \tag{8.27}$$

ou de façon synthétique :

$$\mathbb{V}\left(\overline{Y}_n\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathbb{V}(Y_i) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n}\mathbb{C}ov\left(Y_i,Y_j\right)\right) \tag{8.28}$$

Toutes les variables aléatoires Y_i ont la même variance $\mathbb{V}(Y_i) = \sigma^2$. Par ailleurs, ces variables sont indépendantes, donc toutes les covariances $\mathbb{C}ov\left(Y_i,Y_j\right)$ pour $j \neq i$ sont nulles. Par conséquent, il vient :

$$\mathbb{V}\left(\overline{Y}_n\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(Y_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{n \times \sigma^2}{n^2} = \frac{\sigma^2}{n} \tag{8.29}$$

Ainsi, on obtient :

$$\mathbb{E}\left(\overline{Y}_n\right) = m \tag{8.30}$$

$$\lim_{n\to\infty}\mathbb{V}\left(\overline{Y}_n\right) = \lim_{n\to\infty}\frac{\sigma^2}{n} = 0 \tag{8.31}$$

La moyenne empirique $\overline{Y}_n$ converge en probabilité vers l'espérance m :

$$\overline{Y}_n \xrightarrow{p} m \tag{8.32}$$

■

FOCUS

Loi forte des grands nombres

En imposant des conditions supplémentaires sur les variables Y_i, il est possible d'obtenir une **loi forte des grands nombres** ou **théorème de Kolmogorov**. À la différence de la loi faible, la loi forte implique une convergence presque sûre de la moyenne empirique vers l'espérance :

$$\overline{Y}_n \overset{a.s.}{\to} \mathbb{E}(Y_i) \tag{8.33}$$

Ce qu'il y a de remarquable dans la loi faible des grand nombres, c'est que ce résultat s'applique **quelle que soit la distribution** des variables aléatoires $Y_1,\ldots,Y_n$. La seule hypothèse est que ces variables doivent être indépendantes (et identiquement distribuées, avec les mêmes espérances et variances dans le cas du théorème de Khintchine). Que les variables Y_i aient une distribution de Student, de Poisson, du khi-deux ou une distribution non standard, leur moyenne empirique $\overline{Y}_n$ converge toujours en probabilité vers leur espérance.

Exemple

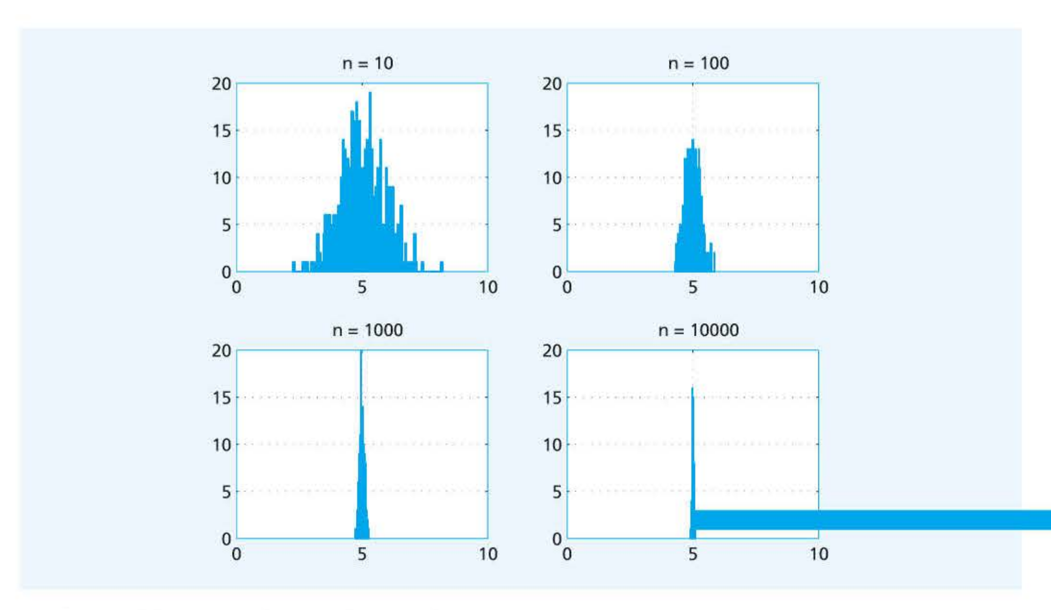
On considère n variables aléatoires discrètes $Z_1,\ldots,Z_n$ indépendantes et identiquement distribuées (i.i.d.) telles que $Z_i \sim \mathcal{P}(\lambda)$. D'après les propriétés de la loi de Poisson (▶ chapitre 7), on sait que $\mathbb{E}(Z_i) = \lambda$. Par conséquent, d'après le théorème de Khintchine, on a :

$$\overline{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i \overset{p}{\to} \lambda \tag{8.34}$$

Afin d'illustrer cette propriété, menons l'expérience suivante. On considère des variables aléatoires indépendantes et identiquement distribuées selon une loi uniforme $Y_i \sim U_{[0,10]}$, $\forall i = 1,\ldots,n$, avec $\mathbb{E}(Y_i) = 5$. On applique la procédure suivante :

1. Grâce à un logiciel, on tire des réalisations $\{y_1,\ldots,y_n\}$ des n variables $\{Y_1,\ldots,Y_n\}$. Si $n = 3$, on obtient par exemple trois réalisations $\{1{,}7363; 4{,}9926; 7{,}6626\}$.
2. On calcule une réalisation de la moyenne empirique $\overline{Y}_n$. Cette réalisation est notée $\overline{y}_n = n^{-1}\sum_{i=1}^{n} y_i$. Dans l'exemple précédent, on obtient $\overline{y}_n = 4{,}8105$.
3. On répète 5 000 fois les étapes 1 et 2. On obtient ainsi 5 000 réalisations de la variable $\overline{Y}_n$.
4. On construit l'histogramme de ces 5 000 réalisations.
5. On répète l'expérience pour différentes valeurs de la dimension n (taille d'échantillon). Il convient de ne pas confondre ici la taille d'échantillon n (par exemple 3) et le nombre de réplications (5 000).

Sur la figure 8.4 sont reportés les histogrammes des 10 000 réalisations $\overline{y}_n$ obtenues pour quatre valeurs de n, à savoir $n = 10$, $n = 100$, $n = 1\,000$ et $n = 10\,000$. On observe que lorsque la dimension n est très grande, les réalisations de la moyenne empirique $\overline{Y}_n$ tendent à se concentrer autour de la valeur de l'espérance $\mathbb{E}(Y_i) = 5$. On aurait pu réaliser la même expérience avec des variables aléatoires Y_i admettant une autre loi (Student, khi-deux, etc.). Le résultat aurait été le même, ce qui confirme le caractère général de la loi faible des grands nombres qui ne dépend pas de la forme de la distribution de Y_i.



▲ **Figure 8.4** Illustration de la loi faible des grands nombres

1.3 Convergence en moyenne quadratique

Dans la pratique, la notion de **convergence en moyenne quadratique** est moins utilisée que la notion de convergence en probabilité ou de convergence presque sûre. Elle est surtout utilisée pour démontrer ces deux dernières.

Définition 8.4

La suite de variables aléatoires (X_n), telle que $\mathbb{E}(|X_n|^2) < \infty$, **converge en moyenne quadratique** vers une constante c, si lorsque n tend vers l'infini :

$$\mathbb{E}(|X_n - c|^2) < \gamma \tag{8.35}$$

pour toute valeur $\gamma > 0$. On note la convergence en moyenne quadratique sous la forme suivante :

$$X_n \overset{m.s.}{\to} c \tag{8.36}$$

Le symbole « m.s. » renvoie à la traduction anglo-saxonne du terme moyenne quadratique (*mean square*). Une façon équivalente de définir la convergence en moyenne quadratique est de poser que $X_n \overset{m.s.}{\to} c$ si :

$$\lim_{n\to\infty} \mathbb{E}(|X_n - c|^2) = 0 \tag{8.37}$$

L'idée est toujours la même : la suite (X_n) converge vers une constante c, si sa distribution est centrée sur c, *i.e.* si $\mathbb{E}(X_n) = c$ et si sa variance $\mathbb{E}((X_n - c)^2)$ tend vers 0 lorsque la dimension n tend vers l'infini. La densité de la variable X_n devient alors extrêmement concentrée autour de la valeur c.

Remarque : La convergence en moyenne quadratique implique la convergence en probabilité. Si $X_n \overset{m.s.}{\to} c$ alors $X_n \overset{p}{\to} c$, la réciproque n'est pas nécessairement vraie. Afin de prouver ce résultat, nous pouvons utiliser l'inégalité de Tchebychev.

Proposition

Inégalité de Tchebychev

Soit X une variable aléatoire telle que $\mathbb{E}(X) = \mu$ existe et soit finie, et de variance égale à $\mathbb{V}(X) = \sigma^2$. Alors, $\forall k \in \mathbb{R}^+$:

$$\Pr(|X - \mu| > k\sigma) \leq \frac{1}{k^2} \tag{8.38}$$

Appliquons l'inégalité de Tchebychev pour démontrer que la convergence en moyenne quadratique implique la convergence en probabilité. Pour cela, il suffit de remarquer que si $X_n \overset{m.s.}{\to} c$ avec $\mathbb{E}(X_n) = c$ alors $\forall \gamma \in \mathbb{R}^+$:

$$\mathbb{E}(\,|X_n - c|^2) = \mathbb{E}(\,(X_n - \mathbb{E}(X_n))^2) = \mathbb{V}(X_n) = \sigma^2 < \gamma \tag{8.39}$$

D'après l'inégalité de Tchebychev, si l'on pose $\delta = k\sigma$, il vient :

$$\Pr(|X - c| > \delta) \leq \frac{\sigma^2}{\delta^2} \tag{8.40}$$

D'après le résultat de l'équation (8.39), puisque $\delta = k\sigma > 0$, on obtient :

$$\Pr(|X - c| > \delta) \leq \frac{\sigma^2}{\delta^2} < \frac{\gamma}{\delta^2} \tag{8.41}$$

En posant $\varepsilon = \gamma/\delta^2 > 0$, on retrouve l'inégalité qui correspond à la définition de la convergence en probabilité, *i.e.* $\Pr(|X - c| > \delta) \leq \varepsilon$. Ainsi, si $X_n \overset{m.s.}{\to} c$ alors $X_n \overset{p}{\to} c$.

1.4 Convergence en loi

La notion de **convergence en loi (ou en distribution)** est fondamentalement différente des trois notions de convergence étudiées précédemment (presque sûre, probabilité et moyenne quadratique). Pour ces trois notions, nous avons vu qu'une séquence de variables aléatoires, indicée par n, converge vers une constante (quantité déterministe), lorsque la dimension n tend vers l'infini :

$$\underbrace{X_n}_{\text{variable aléatoire}} \overset{m.s.\,/\,p\,/\,a.s.}{\to} \underbrace{c}_{\text{constante}} \tag{8.42}$$

Au contraire, dans le cadre de la convergence en loi (notée d), une séquence de variables aléatoires converge vers une **autre variable aléatoire**, ne dépendant pas de la dimension n :

$$\underbrace{X_n}_{\text{variable aléatoire}} \overset{d}{\to} \underbrace{X}_{\text{variable aléatoire}} \tag{8.43}$$

Définition 8.5

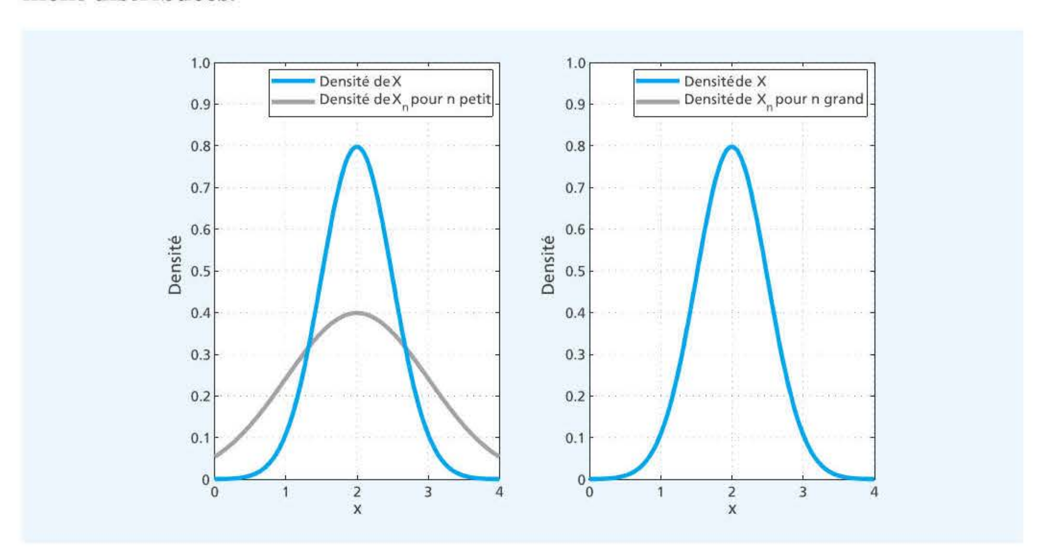
Soit une suite de variables aléatoires (X_n) ayant pour fonction de répartition $F_n(.)$. On dit que (X_n) **converge en loi** (ou en distribution) vers une variable aléatoire X définie sur le support $X(\Omega)$ et ayant pour fonction de répartition $F(.)$ si :

$$\lim_{n\to\infty} F_n(x) = F(x) \quad \forall x \in X(\Omega) \tag{8.44}$$

On note la convergence en distribution sous la forme suivante :

$$X_n \overset{d}{\to} X \tag{8.45}$$

Le symbole « d » renvoie à la traduction anglo-saxonne du terme loi statistique (*distribution*). L'idée de la convergence en loi est la suivante : lorsque n tend vers l'infini, la distribution de la variable X_n est identique à celle d'une autre variable aléatoire, notée X. Leurs fonctions de densité et de répartition sont alors identiques pour toutes les valeurs admissibles sur le support de la loi de X, comme l'illustre la figure 8.5. Dit autrement, lorsque n tend vers l'infini, les variables aléatoires X_n et X sont **identiquement distribuées**.



▲ **Figure 8.5 Illustration de la notion de convergence en loi**

Comme nous l'avions mentionné précédemment, la **convergence en probabilité** implique la **convergence au loi**, dans un sens particulier. En effet, si l'on note :

$$X_n \overset{p}{\to} X \tag{8.46}$$

cela signifie que la suite définie par la différence entre les deux variables aléatoires X_n et X converge (en probabilité) vers 0 :

$$X_n - X \overset{p}{\to} 0 \tag{8.47}$$

Ainsi, lorsque n tend vers l'infini, il n'y a pratiquement pas de différences entre les réalisations de X_n et celles de X. Les deux variables ont donc la même fonction de répartition, ce qui implique que X_n converge en distribution vers X :

$$X_n \overset{d}{\to} X \tag{8.48}$$

Remarque : Au sens strict, la convergence en loi implique la convergence d'une séquence de variables aléatoires (X_n) vers une variable aléatoire X ne dépendant pas de n :

$$\underbrace{X_n}_{\text{variable aléatoire}} \overset{d}{\to} \underbrace{X}_{\text{variable aléatoire}} \tag{8.49}$$

Mais parfois, on note la convergence en loi de la façon suivante :

$$\underbrace{X_n}_{\text{variable aléatoire}} \overset{d}{\to} \underbrace{\mathcal{L}}_{\text{loi asymptotique}} \tag{8.50}$$

où $\mathcal{L}$ désigne la loi de la variable aléatoire X.

Exemple

Supposons que X_n converge en loi vers la variable X, telle que $X \sim \mathcal{N}(0,1)$. On peut alors noter plus simplement que :

$$X_n \xrightarrow{d} \mathcal{N}(0,1) \tag{8.51}$$

Une des principales applications de la notion de convergence en loi est le théorème central limite.

2 Théorème central limite

Le **théorème central limite** permet d'étudier la convergence en loi d'une **transformée de la moyenne empirique** de variables aléatoires indépendantes. C'est sans conteste le théorème fondamental de la statistique mathématique.

Pourquoi s'intéresser spécifiquement au comportement de la moyenne empirique dans le cadre de la théorie de l'estimation ? Comme nous le verrons par la suite, un estimateur est une variable aléatoire définie comme une fonction des variables de l'échantillon. Cette variable (ou séquence de variables aléatoires) dépend donc de la taille de l'échantillon n et il convient d'étudier son comportement asymptotique lorsque n tend vers l'infini.

Quel est le lien avec la moyenne empirique ? Ici réside une des principales « astuces » de la théorie asymptotique : il est généralement possible d'exprimer n'importe quel estimateur comme une moyenne empirique ou comme une fonction de la moyenne empirique des variables aléatoires de l'échantillon. Ainsi, connaissant la convergence en loi d'une transformée de la moyenne empirique *via* le théorème central limite, on est en mesure d'en déduire, sous certaines hypothèses, la distribution asymptotique de la plupart des estimateurs usuels (maximum de vraisemblance, moindres carrés ordinaires, estimateurs de moments, etc.).

Le théorème central limite constitue ainsi le théorème fondamental sur lequel se base l'essentiel de la **théorie de l'estimation** (► chapitre 9) et de la **théorie de l'inférence** (► chapitre 11). C'est dire l'importance de ce théorème pour la suite de cet ouvrage.

2.1 Vitesse de convergence

À ce stade du chapitre, que savons nous concernant le comportement de la moyenne empirique $\overline{Y}_n$ de variables aléatoires réelles $Y_1,\ldots,Y_n$ **indépendantes** ?

$$\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{8.52}$$

Supposons que les variables aléatoires $Y_1,\ldots,Y_n$ soient par ailleurs **identiquement distribuées**, ou à tout le moins, qu'elles aient la même espérance $\mathbb{E}(Y_i) = m$, $\forall i = 1,\ldots,n$, et la même variance $\mathbb{V}(Y_i) = \sigma^2$, $\forall i = 1,\ldots,n$. D'après la loi faible des grands nombres (théorème de Khintchine), nous savons que cette moyenne empirique converge en probabilité vers l'espérance m :

$$\overline{Y}_n \xrightarrow{p} m \tag{8.53}$$

Ce résultat signifie que lorsque n tend vers l'infini, la moyenne empirique tend à être une variable aléatoire « dégénérée » : elle se réduit « presque » à une quantité déterministe égale à m (constante) puisque sa variance tend vers 0. Sous des conditions supplémentaires sur les variables Y_i, la variable $\overline{Y}_n$ peut même converger presque sûrement vers l'espérance m (loi forte des grands nombres ou théorème de Kolmogorov) :

$$\overline{Y}_n \overset{a.s.}{\rightarrow} m \tag{8.54}$$

Dans ce cas, lorsque n tend vers l'infini, la moyenne empirique n'est plus une variable aléatoire. Sa distribution pour $n \rightarrow \infty$ est une **masse ponctuelle** (▶ figure 8.1) et sa fonction de densité devient alors :

$$\lim_{n\rightarrow\infty} f_{\overline{Y}_n}(x) = f(x) \quad \forall x \in \mathbb{R} \tag{8.55}$$

$$f(x) = \begin{cases} 1 & \text{si } x = m \\ 0 & 0 \text{ sinon} \end{cases} \tag{8.56}$$

Ce résultat est problématique : lorsque la dimension n tend vers l'infini, la distribution de la moyenne empirique est dégénérée et il n'est pas possible de construire une théorie de l'inférence à partir de cette distribution.

La solution consiste à **transformer** la variable $\overline{Y}_n$ de sorte à ce que la variable transformée converge en loi vers une **distribution non dégénérée**, c'est-à-dire une distribution dont la variance ne tende ni vers 0, ni vers l'infini (auquel cas la densité serait non définie). Comme nous allons le découvrir dans l'énoncé du théorème central limite, cette transformation est de la forme $\sqrt{n}\left(\overline{Y}_n - m\right)$. Dans cette transformation, l'élément le plus important est le terme $\sqrt{n}$ qui détermine la **vitesse de convergence** de la variable transformée.

Pourquoi multiplier la moyenne empirique par $\sqrt{n}$? Afin de mieux comprendre ce résultat, supposons pour simplifier que $m = \mathbb{E}(Y_i) = 0$ et étudions le comportement asymptotique de la variable transformée $n^{\alpha}\overline{Y}_n$ pour $\alpha \geq 0$. Sous l'hypothèse d'indépendance, il vient immédiatement que :

$$\mathbb{E}\left(n^{\alpha}\overline{Y}_n\right) = n^{\alpha}\mathbb{E}\left(\overline{Y}_n\right) = n^{\alpha}m = 0 \tag{8.57}$$

$$\mathbb{V}\left(n^{\alpha}\overline{Y}_n\right) = n^{2\alpha}\mathbb{V}\left(\overline{Y}_n\right) = n^{2\alpha}\frac{\sigma^2}{n} = n^{2\alpha-1}\sigma^2 \tag{8.58}$$

Considérons trois cas suivant les valeurs du paramètre α :

Premier cas. Si $\alpha > 1/2$, alors $2\alpha - 1 > 0$. La variance de la variable $n^{\alpha}\overline{Y}_n$ tend vers l'infini avec la dimension n, puisque :

$$\lim_{n\rightarrow\infty} \mathbb{V}\left(n^{\alpha}\overline{Y}_n\right) = \sigma^2 \lim_{n\rightarrow\infty} n^{2\alpha-1} = +\infty \quad \text{car } 2\alpha - 1 > 0 \tag{8.59}$$

Deuxième cas. Si $\alpha < 1/2$ alors $2\alpha - 1 < 0$. Dans ce cas, la variance de $n^{\alpha}\overline{Y}_n$ tend vers 0 :

$$\lim_{n\rightarrow\infty} \mathbb{V}\left(n^{\alpha}\overline{Y}_n\right) = \sigma^2 \lim_{n\rightarrow\infty} n^{2\alpha-1} = 0 \quad \text{car } 2\alpha - 1 < 0 \tag{8.60}$$

Puisque $\mathbb{E}\left(n^{\alpha}\overline{Y}_n\right) = 0$, la variable $n^{\alpha}\overline{Y}_n$ converge en probabilité vers $m = 0$.

Troisième cas. Si l'on suppose que le paramètre α est précisément égal à 1/2, alors $2\alpha - 1 = 0$. La variance de la variable $n^{\alpha}\overline{Y}_n$ devient :

$$\mathbb{V}\left(n^{\alpha}\overline{Y}_n\right) = n^{2\alpha-1}\sigma^2 = n^0\sigma^2 = \sigma^2 \tag{8.61}$$

Cette variance est donc indépendante de n. Lorsque n tend vers l'infini, la variance reste égale à σ^2 :

$$\lim_{n\to\infty} \mathbb{V}\left(n^{\alpha}\overline{Y}_n\right) = \sigma^2 \text{ si } \alpha = 1/2 \quad (8.62)$$

Ainsi, seule la variable transformée $\sqrt{n} \times \overline{Y}_n$ (obtenue pour $\alpha = 1/2$) admet une variance non dégénérée lorsque la dimension n tend vers l'infini. De façon générale pour $m \neq 0$, cela signifie que la variable transformée $\sqrt{n}\left(\overline{Y}_n - m\right)$ converge en distribution vers une loi de probabilité non dégénérée. On dit que la variable $\overline{Y}_n - m$ converge à la vitesse $1/\sqrt{n}$, notée $O\left(n^{-1/2}\right)$.

À ce stade, nous connaissons la variance de la variable transformée $\sqrt{n}\left(\overline{Y}_n - m\right)$ lorsque n tend vers l'infini. La question est de savoir quelle est sa distribution. C'est précisément l'objet du théorème central limite.

2.2 Énoncé du théorème central limite

Il existe plusieurs versions du théorème central limite (► focus sur les théorèmes central limite), nous présentons ici celle énoncée par Lindeberg-Levy (1920).

Théorème 8.2

Théorème central limite, Lindeberg-Levy

Soit $Y_1,\ldots,Y_n$ une séquence de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) avec une espérance finie $\mathbb{E}(Y_i) = m$ et une variance finie $\mathbb{V}(Y_i) = \sigma^2$, $\forall i = 1,\ldots,n$. Alors la moyenne empirique $\overline{Y}_n = n^{-1}\sum_{i=1}^{n} Y_i$ vérifie :

$$\sqrt{n}\left(\overline{Y}_n - m\right) \xrightarrow{d} \mathcal{N}\left(0,\sigma^2\right) \quad (8.63)$$

Le principal résultat du théorème central limite est que la *transformée* d'une moyenne empirique de variables i.i.d. converge en distribution vers une **distribution normale**. Il est important de noter que ce n'est pas la moyenne empirique $\overline{Y}_n$ qui converge vers une distribution normale, mais sa transformée $\sqrt{n}\left(\overline{Y}_n - m\right)$. Rappelons que nous avons vu deux résultats concernant la moyenne empirique de variables aléatoires $Y_1,\ldots,Y_n$ indépendantes et identiquement distribuées :

$$\text{Loi faible des grands nombres : } \overline{Y}_n \xrightarrow{p} m \quad (8.64)$$

$$\text{Théorème central limite : } \sqrt{n}\left(\overline{Y}_n - m\right) \xrightarrow{d} \mathcal{N}\left(0,\sigma^2\right) \quad (8.65)$$

Le premier résultat (loi faible des grands nombres) nous donne la convergence en probabilité de la moyenne empirique $\overline{Y}_n$. Ce résultat équivaut à une convergence en loi vers une distribution dégénérée de variance nulle. Le deuxième résultat (théorème central limite) nous donne la convergence en distribution d'une transformée de la moyenne empirique, à savoir $\sqrt{n}\left(\overline{Y}_n - m\right)$. Il ne faut pas confondre les deux résultats.

Remarque : Rappelons que si une variable Z suit une loi normale $\mathcal{N}\left(0,\sigma^2\right)$ alors Z/σ suit une loi $\mathcal{N}(0,1)$ (► chapitre 7). Ainsi, le résultat du théorème central limite peut aussi s'écrire sous la forme :

$$\sqrt{n}\left(\frac{\overline{Y}_n - m}{\sigma}\right) \xrightarrow{d} \mathcal{N}(0,1) \quad (8.66)$$

FOCUS

Les différents théorèmes central limite

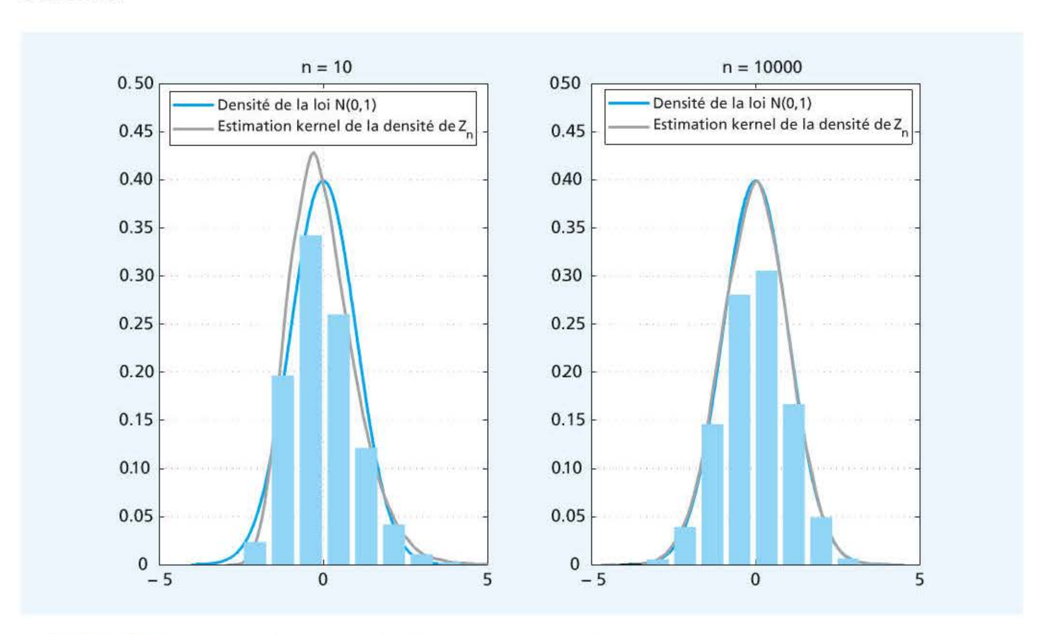
Il n'existe pas *un* théorème central limite, mais *des* théorèmes central limite. Ces théorèmes diffèrent suivant les hypothèses postulées sur les variables $Y_1,\ldots,Y_n$. Le théorème le plus connu est sans doute celui de **Lindeberg-Levy** qui considère une séquence de variables aléatoires indépendantes et identiquement distribuées (i.i.d.). Puisque ces variables ont la même distribution, elles ont la même espérance $\mathbb{E}(Y_i) = m$ et la même variance $\mathbb{V}(Y_i) = \sigma^2, \forall i = 1,\ldots,n$. Une autre version du théorème central limite, celle de **Lindeberg–Feller**, relâche l'hypothèse d'une distribution commune et considère des variances potentiellement différentes, $\mathbb{V}(Y_i) \neq \mathbb{V}\left(Y_j\right)$ pour $i \neq j$. Le théorème central limite de **Lyapounov** s'applique dans le cas de variables ayant des variances et des espérances hétérogènes. Enfin, d'autres versions relâchent l'hypothèse d'indépendance des variables Y_i et considèrent des formes faibles de dépendance.

Comme dans le cas de la loi faible des grands nombres, ce qu'il y a de remarquable dans le théorème central limite, c'est que ce résultat s'applique **quelle que soit la distribution** des variables aléatoires $Y_1,\ldots,Y_n$. La seule hypothèse est que ces variables doivent être indépendantes et identiquement distribuées (dans le cas du théorème de Lindeberg-Levy). Que les variables Y_i aient une distribution de Student, de Poisson, du khi-deux ou une distribution non standard, leur moyenne empirique transformée $\sqrt{n}\left(\overline{Y}_n - m\right)$ converge toujours en distribution vers une loi normale lorsque n tend vers l'infini.

Afin d'illustrer cette propriété, menons l'expérience suivante. On considère des variables aléatoires indépendantes et identiquement distribuées selon une loi du khi-deux $Y_i \sim \chi^2(2)$, $\forall i = 1,\ldots,n$, avec $\mathbb{E}(Y_i) = 2$ et $\mathbb{V}(Y_i) = 4$. On applique la procédure suivante :

1. Grâce à un logiciel, on tire des réalisations $\{y_1,\ldots,y_n\}$ des n variables $\{Y_1,\ldots,Y_n\}$. Si $n = 3$, on obtient par exemple trois réalisations $\{0{,}7444;\ 1{,}5487;\ 1{,}8604\}$.
2. On calcule une réalisation de la moyenne empirique $\overline{Y}_n$. Cette réalisation est notée $\overline{y}_n = n^{-1}\sum_{i=1}^{n} y_i$. Dans l'exemple précédent, on obtient $\overline{y}_n = 1{,}3845$.
3. On considère la variable aléatoire transformée :
$$Z_n = \sqrt{n}\left(\frac{\overline{Y}_n - \mathbb{E}(Y_i)}{\sqrt{\mathbb{V}(Y_i)}}\right) = \sqrt{n}\left(\frac{\overline{Y}_n - 2}{2}\right) \qquad (8.67)$$
À partir de la réalisation de la moyenne empirique $\overline{y}_n$, on calcule une réalisation de cette variable transformée comme $z_n = \sqrt{n}\,(\overline{y}_n - 2)/2$. Dans l'exemple précédent on obtient une valeur de $z_n = -0{,}5331$.
4. On répète cette procédure 5 000 fois (étapes 1 à 3). On obtient alors 5 000 réalisations de la variable z_n.
5. On construit un histogramme de ces 5 000 réalisations et l'on compare cet histogramme à la densité d'une loi normale centrée réduite $\mathcal{N}(0,1)$.

D'après le résultat du théorème central limite, on doit observer que lorsque la dimension n est très grande, la distribution empirique de la variable transformée Z_n doit tendre vers une distribution normale centrée et réduite. Sur la figure 8.6 sont reportés (1) l'histogramme des 5 000 réalisations z_n, (2) la fonction de densité d'une loi normale centrée réduite, *i.e.* $\phi(x) = 1/\sqrt{2\pi}\exp\left(-x^2/2\right)$, et (3) un **estimateur kernel**[2] de la densité de la variable Z_n, obtenus pour deux valeurs de n, à savoir $n = 10$ et $n = 10\,000$. Nous avons inclu un estimateur kernel dans le graphique car l'histogramme n'est pas un bon estimateur de la densité de Z_n. L'estimateur kernel est plus précis et permet de mieux apprécier la convergence de la distribution de Z_n vers la loi normale.



▲ **Figure 8.6** **Illustration du théorème central limite**

On vérifie sur la partie droite de la figure 8.6 que pour $n = 10\,000$, il n'y pratiquement pas de différences entre la distribution empirique de la variable Z_n et la densité d'une loi normale centrée réduite. On pourrait reconduire cette expérience pour n'importe quelle distribution des variables Y_i à la place de la distribution du khi-deux, en utilisant par exemple une loi de Student, une loi de Poisson, etc. Dans tous les cas, la moyenne empirique transformée Z_n converge vers une loi normale, dès lors que les variables Y_i sont indépendantes et identiquement distribuées.

Le théorème central limite peut être étendu au cas multivarié. Supposons que les variables Y_i ne soient plus des scalaires, mais des vecteurs de k variables aléatoires.

$$\underset{(k,1)}{Y_i} = \begin{pmatrix} Y_{i,1} \\ Y_{i,2} \\ \cdots \\ Y_{i,k} \end{pmatrix} \tag{8.68}$$

2 Un estimateur kernel d'une densité est un estimateur non-paramétrique défini comme une sorte de moyenne pondérée des observations de l'échantillon (*cf.* Greene, 2005), les poids étant définis par une fonction dite kernel.

On suppose que l'espérance et la matrice de variance-covariance des variables Y_i sont définies de la façon suivante (▶ chapitre 6) :

$$\underset{(k,1)}{\mathbb{E}(Y_i)} = \begin{pmatrix} \mathbb{E}(Y_{i,1}) \\ \mathbb{E}(Y_{i,2}) \\ \dots \\ \mathbb{E}(Y_{i,k}) \end{pmatrix} = \underset{(k,1)}{\boldsymbol{\mu}} \tag{8.69}$$

$$\underset{(k,k)}{\mathbb{V}(Y_i)} = \begin{pmatrix} \mathbb{V}(Y_{i1}) & \mathbb{C}ov(Y_{i1},Y_{i2}) & \dots & \mathbb{C}ov(Y_{i1},Y_{ik}) \\ \mathbb{C}ov(Y_{i2},Y_{i1}) & \mathbb{V}(Y_{i2}) & \dots & \mathbb{C}ov(Y_{i2},Y_{ik}) \\ \dots & \dots & \dots & \dots \\ \mathbb{C}ov(Y_{ik},Y_{i1}) & \mathbb{C}ov(Y_{ik},Y_{i2}) & \dots & \mathbb{V}(Y_{ik}) \end{pmatrix} = \underset{(k,k)}{\boldsymbol{\Sigma}} \tag{8.70}$$

L'énoncé du théorème central limite multivarié est alors le suivant :

Théorème 8.3

Théorème central limite multivarié

Soit $\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_n$ une séquence de vecteurs de variables aléatoires de dimension $k \times 1$. On suppose que ces vecteurs sont indépendants et identiquement distribués (i.i.d.) avec une espérance finie $\mathbb{E}(\boldsymbol{Y}_i) = \boldsymbol{\mu}$ et une matrice de variance-covariance finie $\mathbb{V}(\boldsymbol{Y}_i) = \boldsymbol{\Sigma}$, $\forall i = 1,\dots,n$. Alors la moyenne empirique $\overline{\boldsymbol{Y}}_n = n^{-1}\sum_{i=1}^{n} \boldsymbol{Y}_i$ vérifie :

$$\underbrace{\sqrt{n}\left(\overline{\boldsymbol{Y}}_n - \boldsymbol{\mu}\right)}_{(k\times 1)} \xrightarrow{d} \mathcal{N}\left(\underbrace{\boldsymbol{0}}_{(k\times 1)}, \underbrace{\boldsymbol{\Sigma}}_{(k\times k)}\right) \tag{8.71}$$

2.3 Distribution asymptotique

Nous pouvons à présent introduire la notion de **distribution asymptotique**, qui peut être définie comme suit.

Définition 8.6

Si la variable aléatoire (séquence) X_n converge en loi vers X ayant pour fonction de répartition $F(.)$, alors $F(.)$ est la fonction de répartition de la **distribution asymptotique de** X_n.

$$X_n \xrightarrow{d} \text{loi asymptotique} \tag{8.72}$$

Exemple

Supposons que X_n converge en loi vers une variable X, telle que $X \sim \chi^2(v)$:

$$X_n \xrightarrow{d} \chi^2(v) \tag{8.73}$$

La distribution du khi-deux à v degrés de liberté est la distribution asymptotique de la variable X_n, obtenue lorsque n tend vers l'infini.

Bien souvent dans la suite de cet ouvrage, nous obtiendrons, grâce au théorème central limite, des résultats sur des estimateurs X_n (variables aléatoires) du type :

$$\sqrt{n}(X_n - m) \xrightarrow{d} \mathcal{N}\left(0,\sigma^2\right) \tag{8.74}$$

Au sens strict, ce résultat ne veut pas dire que la distribution asymptotique de X_n est normale, puisque seule la variable transformée $\sqrt{n}\,(X_n - m)$ converge en distribution vers une loi normale. Ainsi, la distribution asymptotique de la variable transformée $\sqrt{n}\,(X_n - m)$ est une loi normale $\mathcal{N}\left(0,\sigma^2\right)$. La question est de savoir si l'on peut déduire de ce résultat une approximation de la distribution asymptotique de la variable X_n ? On admet que lorsque la dimension n est très grande, mais finie, la distribution de la variable X_n est **approximativement asymptotiquement distribuée** selon une loi normale :

$$X_n \overset{asy}{\approx} \mathcal{N}\left(m, \frac{\sigma^2}{n}\right) \tag{8.75}$$

où le signe $\approx$ signifie « approximativement distribué selon » et l'acronyme *asy* renvoie à la notion d'asymptotique (n grand). En effet, le résultat de l'équation (8.74) implique que pour n très grand :

$$\sqrt{n}\,(X_n - m) \overset{asy}{\approx} \mathcal{N}\left(0,\sigma^2\right) \tag{8.76}$$

On sait que pour $a > 0$, si la variable $a \times Z$ suit une loi normale $\mathcal{N}\left(0,\sigma^2\right)$, alors Z suit une loi normale $\mathcal{N}\left(0,\sigma^2/a^2\right)$ (▸ chapitre 7). Dès lors, si l'on suppose que n est une quantité finie (par exemple $n = 100\,000$), on peut écrire :

$$X_n - m \overset{asy}{\approx} \mathcal{N}\left(0, \frac{\sigma^2}{n}\right) \tag{8.77}$$

De même, si la variable $Z - b$ suit une loi normale $\mathcal{N}\left(0,\sigma^2\right)$, alors Z suit une loi normale $\mathcal{N}\left(b,\sigma^2\right)$. Donc, le résultat de l'équation (8.74) peut être compris comme :

$$X_n \overset{asy}{\approx} \mathcal{N}\left(m, \frac{\sigma^2}{n}\right) \tag{8.78}$$

Définition 8.7

Soit une suite de variables aléatoires (X_n) telle que :

$$\sqrt{n}\,(X_n - m) \overset{d}{\to} \mathcal{N}\left(0,\sigma^2\right) \tag{8.79}$$

Lorsque la dimension n est *finie* et *très grande*, la variable X_n est **(approximativement) asymptotiquement distribuée** selon une loi normale :

$$X_n \overset{asy}{\approx} \mathcal{N}\left(m, \frac{\sigma^2}{n}\right) \tag{8.80}$$

C'est pourquoi dans la plupart des ouvrages, dès lors que l'on a :

$$\sqrt{n}\,(X_n - m) \overset{d}{\to} \mathcal{N}\left(0,\sigma^2\right) \tag{8.81}$$

on dit que la variable X_n est *asymptotiquement normalement distribuée*, même si au sens strict c'est inexact puisque seule la variable $\sqrt{n}\,(X_n - m)$ admet exactement une distribution asymptotique normale.

Définition 8.8

L'**espérance asymptotique** et la **variance asymptotique** de la variable X_n, notées $\mathbb{E}_{asy}(X_n)$ et $\mathbb{V}_{asy}(X_n)$, correspondent à l'espérance et la variance de sa loi asymptotique.

Exemple

Par exemple, si une suite de variables aléatoires (X_n) vérifie :

$$\sqrt{n}(X_n - m) \xrightarrow{d} \mathcal{N}\left(0,\sigma^2\right) \quad (8.82)$$

Cela signifie que la distribution asymptotique de X_n est approximativement normale :

$$X_n \overset{asy}{\approx} \mathcal{N}\left(m,\frac{\sigma^2}{n}\right) \quad (8.83)$$

On en déduit que :

$$\mathbb{E}_{asy}(X_n) = m, \quad \mathbb{V}_{asy}(X_n) = \frac{\sigma^2}{n} \quad (8.84)$$

2.4 Théorème de Slutsky et méthode delta

Une dernière question est de savoir comment à partir du théorème central limite obtenir la distribution de n'importe quel estimateur (variable aléatoire), qui peut s'écrire sous la forme d'une fonction de la moyenne empirique. Pour cela nous allons introduire deux outils : le **théorème de Slutsky** et la **méthode delta**.

Théorème 8.4

Théorème de Slutsky

Soient X_n et Y_n deux séquences de variables aléatoires telles que $X_n \xrightarrow{d} X$ et $Y_n \xrightarrow{p} c$, avec $c \neq 0$, alors :

$$X_n + Y_n \xrightarrow{d} X + c \quad (8.85)$$

$$X_n Y_n \xrightarrow{d} cX \quad (8.86)$$

$$\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c} \quad (8.87)$$

Exemple

Supposons que $X_n \xrightarrow{d} \mathcal{N}\left(m,\sigma^2\right)$ et que $Y_n \xrightarrow{p} 2$, alors la séquence de variables aléatoires définie par le ratio X_n/Y_n converge en distribution :

$$\frac{X_n}{Y_n} \xrightarrow{d} \mathcal{N}\left(\frac{m}{2},\frac{\sigma^2}{4}\right) \quad (8.88)$$

La **méthode delta** permet quant à elle de dériver la distribution asymptotique d'une variable aléatoire qui est une fonction d'une autre variable asymptotiquement normalement distribuée.

Définition 8.9

Méthode delta

Soit Z_n une séquence de variables aléatoires indicée par n telle que :

$$\sqrt{n}\,(Z_n - \mu) \xrightarrow{d} \mathcal{N}\left(0,\sigma^2\right) \qquad (8.89)$$

Si $g\,(.)$ est une fonction continue, continûment différentiable et ne dépendant pas de n, alors :

$$\sqrt{n}\,(g\,(Z_n) - g\,(\mu)) \xrightarrow{d} \mathcal{N}\left(0, \left(\left.\frac{\partial g\,(x)}{\partial x}\right|_{\mu}\right)^2 \sigma^2\right) \qquad (8.90)$$

où $\partial g\,(x)/\partial x|\,\mu$ correspond à la dérivée partielle de la fonction $g\,(x)$ par rapport à x évaluée au point $x = \mu$.

Pour la définition de la méthode delta dans le cas d'une distribution multivariée, on pourra se reporter à Greene (2005). Considérons deux exemples d'application de cette méthode.

Exemple

On considère une suite de variables aléatoires (Y_n) telle que

$$\sqrt{n}Y_n \xrightarrow{d} \mathcal{N}\,(0,1) \qquad (8.91)$$

Déterminons la distribution asymptotique de la séquence $(\exp\,(Y_n))$ à partir de la méthode delta. Pour ce faire on définit une fonction $g\,(x) = \exp\,(x)$. Cette fonction est continue, continûment différentiable et ne dépend pas de n. L'espérance asymptotique de la variable $(\exp\,(Y_n))$ est donc égale à :

$$g\left(\mathbb{E}_{asy}\left(\sqrt{n}Y_n\right)\right) = g\,(0) = \exp\,(0) = 1 \qquad (8.92)$$

Afin de déterminer la variance asymptotique de cette séquence, on doit calculer la dérivée première de la fonction $g\,(x)$

$$\frac{\partial g\,(x)}{\partial x} = \exp\,(x) \qquad (8.93)$$

et l'évaluer au point $x = \mathbb{E}_{asy}\left(\sqrt{n}Y_n\right) = 0$. Il vient :

$$\left(\left.\frac{\partial g\,(x)}{\partial x}\right|_{0}\right) = \exp\,(0) = 1 \qquad (8.94)$$

Dès lors, la variance asymptotique de la séquence $(\exp\,(Y_n))$ est égale à :

$$\left(\left.\frac{\partial g\,(x)}{\partial x}\right|_{0}\right)^2 \times \mathbb{V}_{asy}\left(\sqrt{n}Y_n\right) = 1 \times 1 = 1 \qquad (8.95)$$

Par application de la méthode delta, on obtient finalement la distribution asymptotique de $(\exp\,(Y_n))$:

$$\sqrt{n}\,(g\,(Y_n) - g\,(0)) \xrightarrow{d} \mathcal{N}\,(0,1) \qquad (8.96)$$

ou encore :

$$\sqrt{n}\,(\exp\,(Y_n) - 1) \xrightarrow{d} \mathcal{N}\,(0,1) \qquad (8.97)$$

Exemple

On considère n variables $X_1,\ldots,X_n$ i.i.d. telles que $\mathbb{E}(X_i) = \alpha\beta$ et $\mathbb{V}(X_i) = \alpha\beta^2$, avec $\alpha > 0$, $\beta > 0$. Quelle est alors la distribution asympotique de la variable $\widehat{\beta}$ définie par :

$$\widehat{\beta} = \frac{1}{\alpha n}\sum_{i=1}^{n} X_i \tag{8.98}$$

Pour répondre à cette question, remarquons tout d'abord que cette variable s'exprime en fonction de la moyenne empirique $\overline{X}_n$ des variables X_i :

$$\widehat{\beta} = \frac{\overline{X}_n}{\alpha} \tag{8.99}$$

Sachant que les variables $X_1,\ldots,X_n$ sont i.i.d., on peut appliquer le théorème central limite (Lindeberg-Levy) pour obtenir la distribution asymptotique de $\overline{X}_n$. On obtient immédiatement que :

$$\sqrt{n}\left(\overline{X}_n - \alpha\beta\right) \xrightarrow{d} \mathcal{N}\left(0,\alpha\beta^2\right) \tag{8.100}$$

On définit alors une fonction $g(x) = x/\alpha$. Cette fonction est continue et ne dépend pas de n. Par définition de la variable $\widehat{\beta}$, on a :

$$\widehat{\beta} = \frac{\overline{X}_n}{\alpha} = g\left(\overline{X}_n\right) \tag{8.101}$$

En utilisant la méthode delta, il vient :

$$\sqrt{n}\left(g\left(\overline{X}_n\right) - g(\alpha\beta)\right) \xrightarrow{d} \mathcal{N}\left(0,\left(\left.\frac{\partial g(z)}{\partial z}\right|_{\alpha\beta}\right)^2 \alpha\beta^2\right) \tag{8.102}$$

Où la quantité $\left.\frac{\partial g(z)}{\partial z}\right|_{\alpha\beta}$ correspond à la dérivée $\partial g(z)/\partial z$ évaluée au point $\mathbb{E}\left(\overline{X}_n\right) = \mathbb{E}(X_i) = \alpha\beta$. Dans notre cas, nous avons :

$$g'(z) = \frac{\partial g(z)}{\partial z} = \frac{\partial(z/\alpha)}{\partial z} = \frac{1}{\alpha} \tag{8.103}$$

Donc

$$\left.\frac{\partial g(z)}{\partial z}\right|_{\alpha\beta} = g'(\alpha\beta) = \frac{1}{\alpha} \tag{8.104}$$

On en déduit que :

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0,\frac{\alpha\beta^2}{\alpha^2}\right) \tag{8.105}$$

On obtient au final la distribution de $\widehat{\beta}$:

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0,\frac{\beta^2}{\alpha}\right) \tag{8.106}$$



3 questions à

Andreea **Danci**

Analyste risques, General Electric Money Bank



Quel est votre parcours professionnel et votre mission actuelle chez General Electric Money Bank ?

En 2011, à l'issue de mon master à Orléans, j'ai été embauchée chez General Electric Money Bank. Je travaille actuellement au sein du département de gestion et évaluation des Risques. En tant qu'Analyste Risques Senior, je suis en charge du suivi du risque d'une partie du portefeuille Retail et Corporate. Mes principales responsabilités comprennent notamment l'analyse des modèles de scoring, la construction de nouveaux modèles d'octroi ou de comportement, la validation annuelle (demandée par Federal Reserve) de tous les modèles utilisés en partenariat avec la Global Model Validation Team. Je travaille en outre sur des études statistiques ponctuelles en collaboration avec les équipes des départements finance, comptabilité, pricing et marketing.

Dans le cadre de votre activité, en quoi les notions de convergence et de propriétés asymptotiques vous sont-elles utiles ?

Ces notions sont omniprésentes dans mon activité, même si cela ne se traduit pas nécessairement par des démonstrations. Typiquement, nous utilisons régulièrement de nombreux tests statistiques (tests sur les paramètres des modèles, tests de spécification, tests sur les prévisions). Or, les distributions de la plupart des statistiques de tests sont des distributions asymptotiques. De même, nous faisons implicitement référence à des notions de convergence lorsque nous utilisons de grands échantillons : si le modèle est bien spécifié, l'utilisation d'estimateurs convergents nous permet alors d'obtenir des estimations très précises des probabilités de défauts.

Sous quels types de logiciels travaillez-vous ?

Nous travaillons sur le logiciel SAS en raison de la volumétrie des données que nous avons à manipuler. Ses fonctionnalités sont multiples : utilisation des procédures prédéfinies, automatisation de traitements répétitifs avec le langage « Macro », création de sorties directement exploitables en reporting, etc. ■

Les points **clés**

- Pour une séquence de variables aléatoires, il existe quatre modes possibles de convergence (presque sûre, probabilité, moyenne quadratique et convergence en loi).

- La convergence presque sûre et la convergence en probabilité impliquent que la séquence de variables aléatoires considérée converge vers une constante.

- La loi faible des grands nombres indique que la moyenne empirique de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) converge en probabilité vers l'espérance de ces variables.

- La convergence en loi implique que la variable considérée converge vers une distribution, dite distribution asymptotique.

- Le théorème central limite indique qu'une transformée de la moyenne empirique de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) converge en loi vers une loi normale.

ÉVALUATION

► Corrigés sur www.dunod.com

QCM

Pour chacune des questions suivantes, indiquer si les affirmations sont vraies ou fausses (il peut y avoir plusieurs réponses vraies pour chaque question).

1 Convergence

a. La convergence en loi implique la convergence en probabilité.

b. La convergence en probabilité implique la convergence en loi.

c. La convergence presque sûre implique la convergence en loi.

d. La convergence presque sûre implique la convergence en probabilité.

e. La convergence en probabilité implique la convergence presque sûre.

2 Si $Y_1,\ldots,Y_n$ sont des variables aléatoires indépendantes et identiquement distribuées, alors :

a. La moyenne empirique converge presque sûrement vers l'espérance.

b. La moyenne empirique converge en probabilité vers l'espérance.

c. La moyenne empirique converge en loi vers l'espérance.

d. La moyenne empirique converge en loi vers une loi normale.

e. La moyenne empirique centrée sur l'espérance et multipliée par $\sqrt{n}$ converge en loi vers une loi normale.

3 Si une variable Z_n est asympotiquement normalement distribuée, cela signifie que :

a. La variable Z_n converge en distribution vers une loi normale centrée réduite.

b. La variable Z_n converge en distribution vers une loi normale.

c. La variable $\sqrt{n}\,(Z_n - \mathbb{E}\,(Z_n))$ converge en distribution vers une loi normale.

d. La variable $\sqrt{n}\,(Z_n - \mathbb{E}\,(Z_n))$ converge en distribution vers une loi normale centrée réduite.

e. Pour une dimension n grande et finie, la variable Z_n est approximativement distribuée selon une loi normale.

4 Soit $Y_1,\ldots,Y_n$ une suite de n variables aléatoires et soit $\overline{Y}_n$ la moyenne empirique :

a. Le théorème central limite s'applique si les variables Y_i sont i.i.d.

b. Le théorème central limite s'applique si les variables Y_i sont indépendantes mais avec des espérances différentes.

c. Le théorème central limite s'applique si les variables Y_i sont indépendantes mais avec des variances différentes.

d. Le théorème central limite s'applique si les variables Y_i sont indépendantes mais avec des espérances et des variances différentes.

e. Le théorème central limite s'applique si les variables Y_i sont dépendantes mais identiquement distribuées.

Exercice

5 Convergences

Soient deux variables aléatoires réelles X_1 et X_2 indépendantes et distribuées chacune selon une loi $\mathcal{N}\left(0,\sigma^2\right)$. On considère la variable transformée Y définie par la relation :

$$Y = \sqrt{X_1^2 + X_2^2} \tag{8.107}$$

On admet que cette variable Y suit une loi de Rayleigh avec pour fonction de densité :

$$f_Y\left(y;\sigma^2\right) = \frac{y}{\sigma^2}\exp\left(-\frac{y^2}{2\sigma^2}\right) \quad \forall y \in [0,+\infty[\tag{8.108}$$

1. Quelle est la loi de la variable Y^2/σ^2 ? En déduire la valeur de $\mathbb{E}\left(Y^2\right)$ et de $\mathbb{V}\left(Y^2\right)$.
2. Soit $Y_1,\ldots,Y_n$ une suite de variables aléatoires i.i.d. de même loi que Y. On considère une variable transformée définie par :

$$\widehat{\sigma}^2 = \frac{1}{2n}\sum_{i=1}^{n} Y_i^2 \qquad (8.109)$$

Montrez que cette suite de variables aléatoires converge *en probabilité* vers σ^2.
3. On admet que les variables $Y_1^2,\ldots,Y_n^2$ sont i.i.d. de même loi que Y^2. Montrez que la variable $\widehat{\sigma}^2$ est asymptotiquement normalement distribuée.

Sujets d'examen

6 Théorème central limite (Bibmath.net)

Un fournisseur d'accès à Internet met en place un point local d'accès, qui dessert 5 000 abonnés. À un instant donné, chaque abonné a une probabilité égale à 20 % d'être connecté. Les comportements des abonnés sont supposés indépendants les uns des autres.

1. On note X la variable aléatoire égale au nombre d'abonnés connectés à un instant t. Quelle est la loi de X ? Quelle est son espérance, son écart-type ?
2. On pose :

$$Y = \frac{X - 1\,000}{\sqrt{800}} \qquad (8.110)$$

Justifier précisément que l'on peut approcher la loi de Y par la loi normale $\mathcal{N}(0,1)$.
3. Le fournisseur d'accès souhaite savoir combien de connexions simultanées le point d'accès doit pouvoir gérer pour que sa probabilité d'être saturé à un instant donné soit inférieure à 2,5 %. En utilisant l'approximation précédente, proposer une valeur approchée de ce nombre de connexions.

7 Convergences (Université d'Assas)

Soit (X_n) une suite de variables aléatoires discrètes telles que $\forall n \geq 2$:

$$\Pr(X_n = -n) = \Pr(X_n = n) = \frac{1}{2n^2} \qquad (8.111)$$

$$\Pr(X_n = 0) = 1 - \frac{1}{n^2} \qquad (8.112)$$

1. Déterminer la limite en probabilité de la suite (X_n).
2. La suite (X_n) converge-t-elle en moyenne quadratique vers cette même limite ?

8 Théorème central limite (Bibmath.net)

Il arrive assez souvent que le nombre de réservations pour une liaison aérienne soit supérieur au nombre de passagers se présentant effectivement le jour du vol. Cela est dû à des empêchements imprévisibles de certains passagers et à une politique systématique de certains d'entre eux qui réservent des places sur plusieurs vols de façon à choisir au dernier moment celui qui leur convient le mieux (en raison de la concurrence, et selon les tarifs choisis, les compagnies ne pénalisent pas les clients qui se désistent et ne font payer effectivement que ceux qui embarquent).

Pour compenser ce phénomène, une compagnie aérienne exploitant un avion de 300 places décide de faire de la surréservation (surbooking) en prenant pour chaque vol un nombre $n > 300$ de réservations. S'il se présente plus de 300 passagers à l'embarquement, les 300 premiers arrivés prennent leur vol et les autres sont dédommagés financièrement.

1. On considère que les passagers sont mutuellement indépendants et que la probabilité de désistement de chacun d'eux est égale à 10 %. On note n le nombre de réservations prises par la compagnie pour un vol donné et S_n le nombre (aléatoire) de passagers se présentant à l'embarquement pour ce vol. Donner la loi de S_n, $\mathbb{E}(S_n)$ et $\mathbb{V}(S_n)$.
2. Le directeur commercial de la compagnie aimerait connaitre la valeur maximale de n telle que

$$\Pr(S_n \leq 300) \geq 0{,}99 \qquad (8.113)$$

En utilisant le théorème central limite, proposer une solution approchée de ce problème.

Partie 3

Statistique mathématique

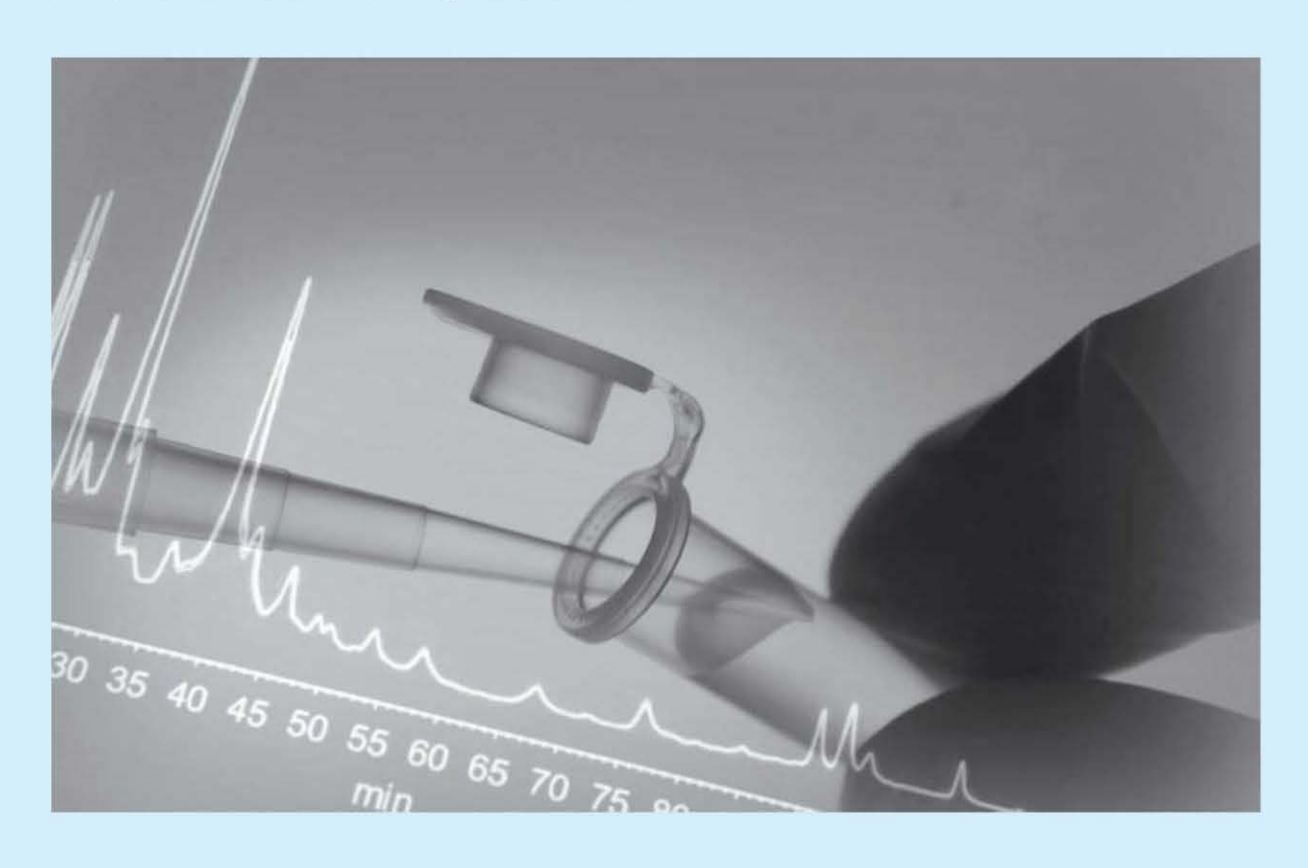
On oppose généralement la statistique descriptive (► partie 1), dont l'objectif est de décrire une réalité statistique (typiquement un échantillon ou une population), à la statistique mathématique dont l'objectif est de modéliser cette réalité et d'apporter des outils d'aide à la décision.

La statistique mathématique est fondée sur deux piliers : la théorie de l'estimation et la théorie des tests ou inférence. Une des méthodes d'estimation les plus utilisées est la méthode dite du maximum de vraisemblance qui peut être appliquée à l'estimation de paramètres de modèles linéaires ou non linéaires.

Chapitre 9

Un laboratoire pharmaceutique souhaite étudier l'effet thérapeutique d'un nouveau médicament sur une population cible. Il réalise pour cela une étude clinique selon un protocole précis auprès d'un échantillon d'individus issus de cette population. Pour tous les individus de cet échantillon, on mesure alors différentes variables d'intérêt (réaction au médicament, effets secondaires, etc.) et l'on peut ainsi calculer des statistiques descriptives sur cet échantillon. Mais l'objectif final n'est pas de décrire les effets du médicament sur les individus de l'échantillon : il s'agit d'évaluer les effets potentiels du médicament sur les individus de la population cible dans son ensemble.

La question générale qui se pose est alors de savoir comment passer d'une information portant sur un échantillon à une information portant sur la population ? C'est précisément l'objet de la **théorie de l'estimation**.



Estimation

Plan

Pré-requis

→ **Connaître** les différentes notions de convergence (▶ chapitre 8).
→ **Connaître** la loi faible des grands nombres.
→ **Connaître** le théorème central limite.

Objectifs

→ **Comprendre** la notion d'échantillon aléatoire.
→ **Comprendre** la notion d'estimateur.
→ **Savoir étudier** les propriétés à distance finie d'un estimateur.
→ **Savoir étudier** les propriétés asymptotiques d'un estimateur.

L'objectif d'une **estimation** est de révéler de l'information sur une caractéristique de la population à partir d'un **échantillon**. Nous nous limiterons dans ce chapitre à la présentation des approches paramétriques[1] de l'estimation : dans ce cadre, on cherche à révéler la valeur d'un ou de plusieurs **paramètres**, associés à la distribution de la caractéristique d'intérêt dans la population. On construit pour cela un **estimateur**. Un estimateur est une variable aléatoire, définie comme une fonction des variables de l'échantillon.

La démarche du statisticien est alors la suivante : on commence par étudier les **propriétés de l'estimateur**. Cela revient à analyser certaines caractéristiques de sa distribution : son espérance, sa variance etc. L'idée générale est de vérifier théoriquement si les réalisations de cette variable aléatoire ont de grandes chances d'être « proches » de la vraie valeur du paramètre que l'on souhaite estimer. On peut aussi comparer différents estimateurs afin de choisir le plus performant : on introduit pour cela les notions d'estimateur optimal et d'estimateur efficace. Une fois que l'on dispose d'un « bon » estimateur, on l'utilise pour obtenir une **estimation**. Une estimation ponctuelle n'est rien d'autre que la réalisation de l'estimateur obtenue à partir de la réalisation de l'échantillon, c'est-à-dire à partir des données statistiques ou des observations. Pour obtenir une estimation, il suffit donc d'appliquer sur les données la « formule » qui définit l'estimateur en fonction des variables de l'échantillon. Cette phase est généralement réalisée à l'aide d'un logiciel d'économétrie ou d'un tableur. Il est aussi possible de fournir un **intervalle de confiance**, c'est-à-dire un encadrement sur la valeur du paramètre que l'on souhaite estimer. Cet encadrement permet de rendre compte de l'incertitude autour de la prévision ponctuelle.

Cette démarche de l'estimation se situe au cœur de très nombreux domaines d'application dans la vie courante et la vie des entreprises : sondages politiques, enquêtes d'opinion, enquêtes économiques, méthodes de scoring, analyses marketing quantitatives, modèles de prévision, etc. Avant d'aborder certains de ces exemples, nous allons à présent introduire les notions d'échantillon et d'estimateur, avant de nous intéresser aux propriétés attendues des estimateurs.

1 Échantillonnage et échantillon

L'objectif de cette section est de présenter le concept d'**échantillon aléatoire**. Ce concept est particulièrement important car il fonde la **théorie de l'estimation**.

Le problème général est le suivant : on souhaite étudier une caractéristique (appelée aussi caractère ou variable statistique) associée à des individus appartenant à une population (▸ chapitre 1). Rappelons qu'une population est un ensemble, fini ou non, d'éléments que l'on souhaite étudier. Il peut s'agir par exemple d'êtres humains (adultes, enfants, chômeurs, salariés, etc.), d'animaux ou encore d'objets (entreprises, voitures, ordinateurs, incendies, accidents, etc.). La caractéristique étudiée peut être qualitative (par exemple la catégorie socio-professionnelle de l'individu : cadre, employé, etc.) ou quantitative (par exemple la taille ou le salaire de l'individu).

1 On oppose les méthodes d'estimation paramétriques aux méthodes non-paramétriques ou semi-paramétriques. Les méthodes paramétriques reposent sur l'hypothèse d'une distribution caractérisée par un nombre fini de paramètres (par exemple une loi normale).

Considérons un exemple simple : supposons qu'une entreprise de l'industrie textile souhaite étudier le poids et la taille (caractéristiques quantitatives) des français et françaises de plus de 18 ans (population de taille finie), afin d'ajuster au mieux ses produits à la morphologie de ses clients. Pour mener à bien cette étude, l'entreprise a deux solutions : le **recensement** ou l'**échantillonnage**.

Définition 9.1

Un **recensement** consiste à mesurer, ou observer, la (ou les) caractéristique(s) d'intérêt de façon exhaustive pour tous les individus de la population.

Une telle solution n'est bien évidemment applicable que lorsque la taille de la population étudiée est relativement faible. Ainsi, à l'époque d'Adam et Eve, un recensement reviendrait, dans notre cas, à peser et à mesurer ces deux individus. Avec deux couples de mesures (80 kg/1,80 m et 55 kg/1,60 m par exemple), on obtiendrait une information complète sur le poids et la taille de la population. Toute méthode d'estimation et de test statistique (inférence) serait alors inutile. Mais aujourd'hui, si l'on admet qu'il y a près de quarante millions de français et de françaises de plus de 18 ans, on imagine facilement que le recensement est de fait impossible[2] pour de nombreuses entreprises : le coût est beaucoup trop élevé. Dans la plupart des cas, il est nécessaire de recourir à la seconde solution : l'**échantillonnage**. L'échantillonnage se définit comme la méthode de construction d'un **échantillon**.

Définition 9.2

Au sens strict, un **échantillon** est un sous-ensemble de la population.

Reprenons notre exemple. Au sens strict, un échantillon consiste en une collection d'individus sélectionnés dans la population française de plus de 18 ans. Le nombre d'individus sélectionnés dans l'échantillon correspond à la **taille de l'échantillon**, notée n. On parle alors de n-échantillon.

Quel est l'intérêt de constituer un échantillon ? L'idée est d'étudier les caractéristiques d'intérêt (poids et taille dans notre cas) pour les individus sélectionnés dans l'échantillon afin d'en tirer de l'information sur ces mêmes caractéristiques pour l'ensemble de la population. Par conséquent, d'un côté la dimension n de l'échantillon doit être suffisamment importante pour que l'on puisse obtenir une information fiable sur la population, mais d'un autre côté elle doit être la plus petite possible afin de limiter le coût de l'enquête.

Une question se pose à ce stade : comment choisir les individus qui composent l'échantillon ? On distingue deux grandes méthodes d'échantillonnage. La première repose sur un choix déterministe des individus. On parle dans ce cas d'**échantillon déterministe** (ou certain) : les individus de l'échantillon ne sont pas choisis « au hasard ». Un exemple lugubre est celui de la décimation[3]. Supposons que tous les individus de

2 Le seul exemple de recensement en France est celui mené de façon régulière par l'Institut national de la statistique et des études économiques (INSEE). Il s'agit d'une enquête portant sur différentes caractéristiques socio-économiques de la population de la France.

3 La décimation était un châtiment appliqué dans l'armée romaine, qui visait à punir les soldats appartenant à une unité s'étant mal conduite au combat. Un soldat sur dix de cette unité était alors mis à mort.

la population soient numérotés de 1 à N, où N désigne la taille de la population. On sélectionne alors de façon systématique tous les individus portant les numéros 1, 10, 20, 30, etc. D'autres exemples d'échantillons certains reposent sur une stratification de la population : on « découpe » la population suivant un grand nombre de caractéristiques, autres que les caractéristiques d'intérêt. On répartit, par exemple, la population française suivant différents critères socio-économiques (catégorie socio-professionnelle, âge, nombre d'enfants, lieu de résidence, etc.). On cherche ensuite à reproduire exactement les mêmes proportions sur ces différents critères dans l'échantillon. On parle alors d'**échantillon par stratification** ou d'**échantillon représentatif**.

Mais en pratique, la méthode la plus utilisée est celle de l'échantillonnage aléatoire : on constitue dans ce cas un **échantillon aléatoire**.

Définition 9.3

Un **échantillon aléatoire** est un échantillon dont les individus sont tirés au hasard parmi la population.

Le tirage de l'échantillon peut se faire avec remise (un même individu de la population peut apparaître plusieurs fois dans l'échantillon) ou sans remise (chaque individu de la population ne peut apparaître qu'une seule fois dans l'échantillon).

Le point essentiel de la notion d'échantillon aléatoire est que les caractéristiques associées aux individus de l'échantillon sont, du fait du tirage au sort des individus, des **variables aléatoires**. Jusqu'à présent, nous n'avons pas évoqué la nature stochastique (aléatoire) ou déterministe (constante) des caractéristiques d'intérêt. Il est parfois compliqué de répondre à cette question, puisqu'il s'agit presque d'un débat philosophique qui renvoie à la vision prédéterminée ou non que l'on se fait du monde. Mais dans le cas de notre exemple, ce statut est clair : à une date donnée, on peut supposer que le poids et la taille d'un individu de la population française sont des quantités déterministes (certaines). Notons x le poids d'un individu, supposé déterministe, et imaginons que notre population soit constituée de quatre individus ($N = 4$) : Pierre, Paul, Jacques et Jean. On suppose que leurs poids exprimés en kilogrammes sont respectivement égaux à :

$$x_{Pierre} = 65 \qquad x_{Jean} = 73 \qquad x_{Paul} = 82 \qquad x_{Jacques} = 68$$

Si l'on souhaite constituer un échantillon aléatoire de taille $n = 2$ (sans remise), il convient de tirer deux individus parmi les quatre individus de la population et d'observer leur poids. Ainsi, on peut obtenir une **réalisation** de l'échantillon du type :

$$(x_{Pierre}, x_{Jean}) = (65{,}73)$$

Mais l'on peut aussi bien obtenir une réalisation du type :

$$\left(x_{Jean}, x_{Jacques}\right) = (73{,}68)$$

ou encore

$$\left(x_{Paul}, x_{Jacques}\right) = (82{,}68)$$

Ainsi, les valeurs observées pour les poids des deux individus de l'échantillon sont **aléatoires** : on peut obtenir $(65{,}73)$, $(73{,}68)$, $(82{,}68)$ ou toute autre combinaison des

valeurs (65, 73, 82, 68). Même si le poids des individus de la population est supposé déterministe (certain), le poids du premier et du deuxième individu de l'échantillon sont des variables aléatoires, tout simplement parce qu'avant le tirage de l'échantillon, on ne sait pas qui seront ces deux individus sélectionnés. Notons X_1 le poids du premier individu de l'échantillon et X_2 le poids du deuxième. Un échantillon aléatoire avec $n = 2$ s'écrit donc sous la forme :

$$(X_1,X_2) \quad (9.1)$$

Définition 9.4

Au sens large, un **n-échantillon aléatoire** est une collection (ou une suite) de variables aléatoires, noté :

$$(X_1,\ldots,X_n) \quad (9.2)$$

où X_i désigne la valeur de la caractéristique d'intérêt associée au $i^{ème}$ individu sélectionné au hasard parmi la population pour constituer l'échantillon.

Remarque : Attention, il convient de ne pas confondre l'*échantillon aléatoire* (collection de variables aléatoires indiquées par une lettre majuscule) et la *réalisation* de cet échantillon (notée avec des lettres minuscules) :

$$\text{Échantillon : } (X_1,\ldots,X_n) \quad (9.3)$$

$$\text{Réalisation (observations) : } (x_1,\ldots,x_n) \quad (9.4)$$

Dans notre exemple, $(x_1,x_2) = \left(x_{Paul},x_{Jacques}\right) = (82,68)$ est une réalisation particulière de l'échantillon aléatoire (X_1,X_2).

2 Estimateur

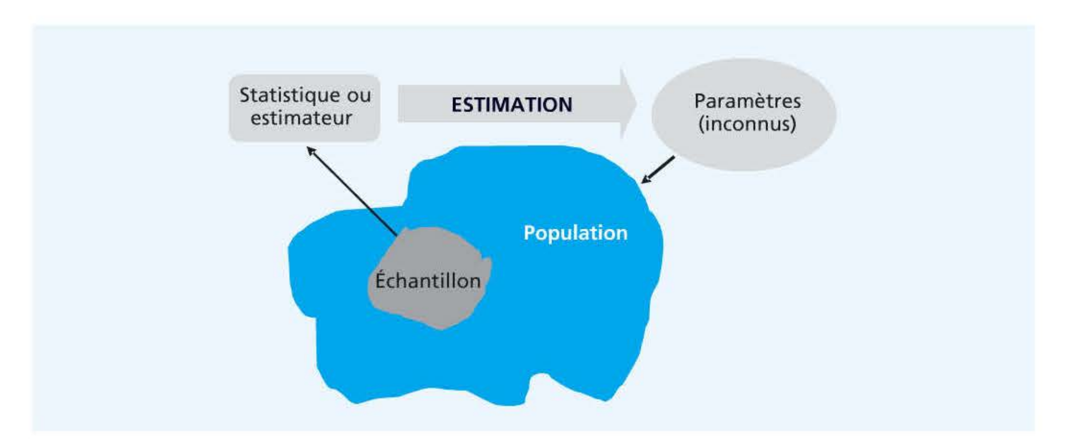
2.1 Principe général de l'estimation

Comme l'indique la figure 9.1, l'objectif d'une procédure d'**estimation** est de révéler de l'information sur le (ou les) paramètre(s) d'intérêt de la population à partir d'un **échantillon** aléatoire. Le problème général est le suivant. On suppose que la caractéristique d'intérêt dans la population, notée X, est une variable aléatoire[4] définie sur un univers probabilisé $(X(\Omega), \mathcal{F}, \Pr)$. La loi de probabilité de cette variable aléatoire est représentée, soit par une *fonction de densité* si X est une variable continue, soit par une *fonction de masse* si X est une variable discrète. On suppose que cette fonction de densité ou de masse dépend d'un **paramètre** θ, qui est *a priori* inconnu et que l'on cherche à estimer. Soit $f_X(x;\theta)$, $\forall x \in X(\Omega)$ la fonction de densité ou de masse de la variable X.

Pour estimer le paramètre θ, on dispose d'un n-échantillon $(X_1,\ldots,X_n)$ dans lequel toutes les variables aléatoires X_i, pour $i = 1,\ldots,n$, sont supposées indépendantes

4 Dans l'exemple de la section précédente, nous avions supposé que les caractéristiques d'intérêt (le poids et la taille en l'occurrence) étaient des variables non stochastiques. En général, on suppose au contraire que la variable statistique étudiée est aléatoire dans la population. Par conséquent, tout échantillon, même issu d'une méthode de sélection déterministe, est une collection de variables aléatoires.

et identiquement distribuées (i.i.d.), de même loi que X. On note $(x_1, \ldots, x_n)$ la réalisation de cet échantillon : cette réalisation correspond aux données (fichier Excel, tableau de valeurs, etc.) utilisées pour l'estimation.



▲ Figure 9.1 Principe général de l'estimation

Exemple

On suppose que la durée de vie d'un équipement, notée D, peut être représentée par une variable aléatoire positive, admettant une distribution exponentielle de paramètre $\lambda > 0$. Ce paramètre est inconnu. Afin de l'estimer, on dispose de six relevés pour lesquels on a pu observer la durée écoulée (exprimée en heures) avant la rupture de l'équipement : (100,102,95,78,135,98). Ces six valeurs correspondent à la réalisation d'un échantillon aléatoire de taille $n = 6$, noté $(D_1,\ldots,D_6)$, où les variables D_i pour $i = 1,\ldots,n$, sont i.i.d. de même loi que D (loi exponentielle).

2.2 Un estimateur est une variable aléatoire

La théorie générale de l'estimation repose sur la notion d'estimateur.

Définition 9.5

Un **estimateur** du paramètre θ est une fonction des variables aléatoires $X_1,\ldots,X_n$ de l'échantillon. Cet estimateur, noté $\widehat{\theta}$, est défini par :

$$\widehat{\theta} = g(X_1,\ldots,X_n) \tag{9.5}$$

Bien évidemment, cette fonction ou cette « formule » $g(.)$ n'est pas choisie au hasard. L'idée est de trouver une fonction qui combine les réalisations de l'échantillon de sorte à révéler de l'information sur le paramètre d'intérêt θ. Nous verrons comment déduire cette fonction, c'est-à-dire comment construire un estimateur, dans la sous-section 2.3 consacrée aux **méthodes d'estimation**. Mais à ce stade, considérons quelques exemples d'estimateurs.

Exemple

Supposons que les variables aléatoires $(Y_1,\ldots,Y_n)$ soient i.i.d. de même loi que Y, où $Y \sim \mathcal{N}\left(m,\sigma^2\right)$. La moyenne empirique (statistique descriptive) :

$$\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{9.6}$$

est un estimateur (de l'espérance m). En effet, $\overline{Y}_n$ est une fonction des variables $Y_1, Y_2, \ldots, Y_n$, telle que :

$$\overline{Y}_n = \frac{1}{n}(Y_1 + \ldots + Y_n) = g(Y_1,\ldots,Y_n) \tag{9.7}$$

Exemple

Supposons que les variables aléatoires $(Y_1,\ldots,Y_n)$ soient i.i.d. de même loi que Y, avec $\mathbb{E}(Y) = m$ et $\mathbb{V}(Y) = \sigma^2$. La variance empirique corrigée :

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \overline{Y}_n\right)^2 \tag{9.8}$$

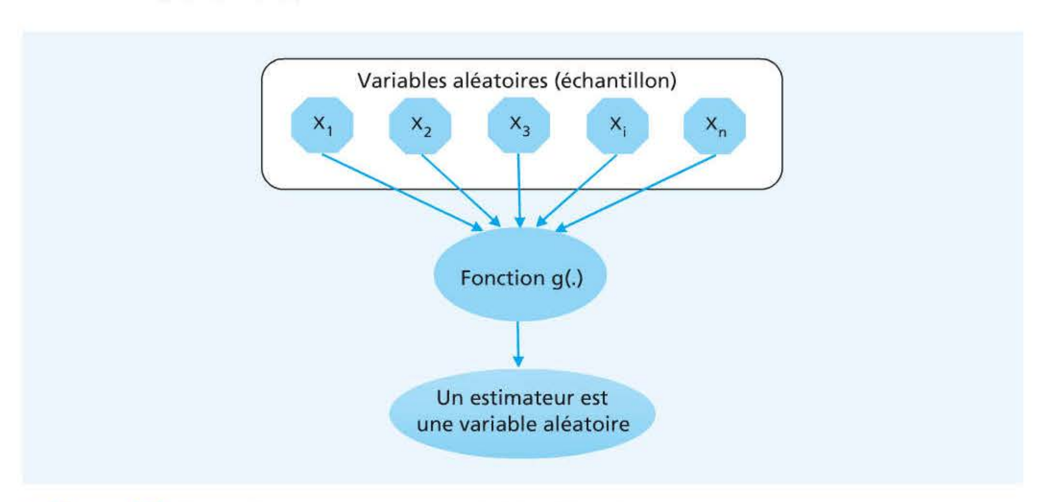
est un estimateur (de la variance σ^2).

Les deux exemples précédents illustrent le fait que toute **statistique** descriptive de l'échantillon (► chapitre 1) est un estimateur, puisque ces statistiques sont généralement définies comme des fonctions (somme, produit, etc.) des variables aléatoires de l'échantillon. Toutefois, il est possible de définir des estimateurs qui ne sont pas des statistiques descriptives usuelles.

Propriété

Un estimateur est une **variable aléatoire**, puisque c'est une fonction des variables aléatoires de l'échantillon.

Si l'on introduit des oranges dans un mixeur, on obtient généralement du jus d'orange. Il en va de même pour les variables aléatoires. Comme l'illustre la figure 9.2, si l'on introduit les variables aléatoires $X_1,\ldots,X_n$ du n-échantillon dans une fonction (somme, produit, etc.), il en ressort une variable aléatoire. C'est pourquoi un estimateur $\widehat{\theta} = g(X_1,\ldots,X_n)$ est une variable aléatoire.



▲ **Figure 9.2 Un estimateur est une variable aléatoire**

Quelle est l'implication de cette propriété ? Si l'estimateur $\widehat{\theta}$ est une variable aléatoire (continue ou discrète)[5], elle est nécessairement caractérisée par une fonction de distribution (fonction de densité dans le cas continu ou fonction de masse dans le cas discret).

Définition 9.6

La distribution de probabilité d'un estimateur (ou d'une statistique) est appelée **distribution d'échantillonnage**.

Exemple

Soit un échantillon (X_1,X_2) de variables i.i.d. telles que $X_i \sim \mathcal{N}(\mu,\sigma^2)$ pour $i = 1,2$. On admet que $\widehat{\mu} = (X_1 + X_2)/2$ est un estimateur du paramètre μ. La loi exacte de l'estimateur $\widehat{\mu}$ est alors la suivante :

$$\widehat{\mu} = \frac{X_1 + X_2}{2} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{2}\right) \quad (9.9)$$

En effet, nous savons que la somme de deux variables normales indépendantes suit une loi normale. Par ailleurs,

$$\mathbb{E}(\widehat{\mu}) = \mathbb{E}\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{2}(\mathbb{E}(X_1) + \mathbb{E}(X_2)) = \frac{2\mu}{2} = \mu \quad (9.10)$$

$$\mathbb{V}(\widehat{\mu}) = \mathbb{V}\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{4}(\mathbb{V}(X_1) + \mathbb{V}(X_2)) = \frac{2\sigma^2}{4} = \frac{\sigma^2}{2} \quad (9.11)$$

puisque les variables X_1 et X_2 sont indépendantes et que leur covariance est nulle.

Comme pour toute variable aléatoire (▸ chapitre 6), on doit distinguer la variable aléatoire elle-même, de sa réalisation. Cette réalisation correspond à une estimation.

Définition 9.7

Une réalisation de l'estimateur $\widehat{\theta}$ associée à une réalisation $(x_1,\ldots,x_n)$ de l'échantillon correspond à une **estimation (ponctuelle)** du paramètre θ. L'estimation est généralement notée $\widehat{\theta}(x)$ pour la différencier de la variable aléatoire (estimateur) $\widehat{\theta}$:

$$\widehat{\theta}(x) = g(x_1,\ldots,x_n) \quad (9.12)$$

Une estimation[6] n'est donc rien d'autre que l'application de la « formule » $g(X_1,\ldots,X_n)$ aux données, c'est-à-dire aux réalisations de l'échantillon $(x_1,\ldots,x_n)$. Reprenons l'exemple précédent.

Exemple

Soit un échantillon (X_1,X_2) de variables i.i.d. telles que $X_i \sim \mathcal{N}(\mu,\sigma^2)$ pour $i = 1,2$. On admet que la variable

$$\widehat{\mu} = \frac{X_1 + X_2}{2} \quad (9.13)$$

5 En règle générale, les estimateurs sont des variables aléatoires continues car les paramètres θ sont définis sur des parties de $\mathbb{R}$.

6 Nous verrons dans la section 5 qu'il existe plusieurs concepts d'estimation (estimation ponctuelle, par intervalle de confiance et par densité). Par défaut, lorsque rien n'est précisé, une estimation correspond à une estimation ponctuelle : il s'agit tout simplement de la réalisation de l'estimateur obtenue sur un échantillon particulier.

est un estimateur de l'espérance μ. Pour une réalisation $(x_1,x_2) = (10,4)$ de l'échantillon, on obtient une estimation (ponctuelle) du paramètre μ égale à :

$$\widehat{\mu}(x) = \frac{10+4}{2} = 7 \tag{9.14}$$

Ainsi à ce stade du chapitre, il convient de bien distinguer la notion d'**estimateur** de la notion d'**estimation** (réalisation) :

$$\text{Estimateur (variable aléatoire)} : \widehat{\theta} \tag{9.15}$$

$$\text{Estimation (constante)} : \widehat{\theta}(x) \tag{9.16}$$

2.3 Méthodes d'estimation

On peut concevoir une méthode d'estimation comme une sorte de recette de cuisine qui permet d'obtenir un estimateur $\widehat{\theta}$ à partir des ingrédients $X_1,\ldots,X_n$. Plus formellement, on définit une méthode d'estimation de la façon suivante.

Définition 9.8

Une **méthode d'estimation** est une méthode mathématique qui permet de dériver la forme fonctionnelle d'un estimateur $\widehat{\theta} = g(X_1,\ldots,X_n)$ à partir des variables aléatoires de l'échantillon $X_1,\ldots,X_n$.

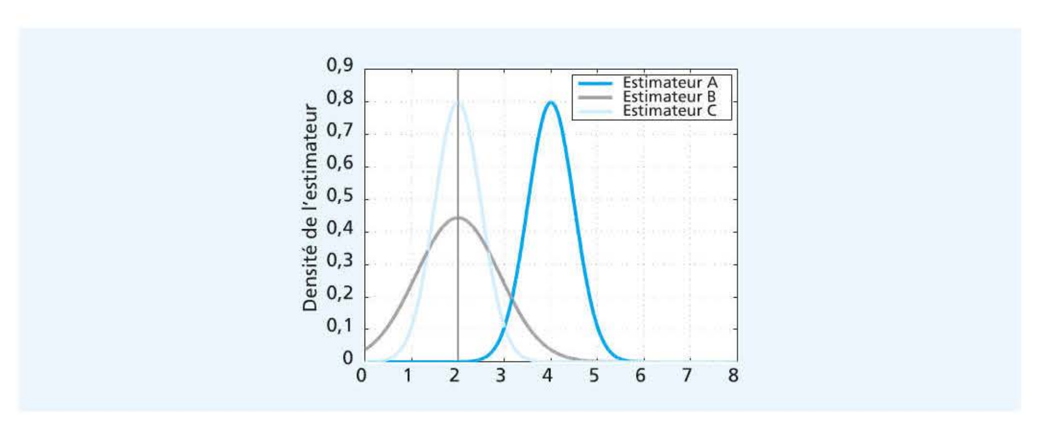
Pour un même problème, on peut parfois appliquer plusieurs méthodes d'estimation. À chaque méthode d'estimation correspond un estimateur particulier. Si l'on se restreint aux seules méthodes d'estimation paramétriques, il existe de nombreuses méthodes suivant le problème étudié et les hypothèses retenues. Citons par exemple :

– la méthode des moindres carrés ordinaires (▶ chapitre 2) ;
– la méthode des moindres carrés généralisés ;
– la méthode du maximum de vraisemblance (▶ chapitre 10) ;
– la méthode des moments généralisés ;
– la méthode des variables instrumentales ;
– la méthode des doubles moindres carrés ordinaires.

2.4 Propriétés d'un estimateur

La question est de savoir ce qu'est un « bon » estimateur. Quelles propriétés doit satisfaire un estimateur pour être considéré comme performant ? Pour répondre à ces questions, nous allons nous concentrer sur la distribution d'échantillonnage de l'estimateur. En étudiant cette distribution, on cherche à vérifier **théoriquement** si l'estimation (réalisation de l'estimateur) a de fortes chances d'être « proche » de la vraie valeur (inconnue) du paramètre θ que l'on souhaite estimer.

Dans l'illustration de la figure 9.3, sont représentées les fonctions de densité de trois estimateurs, notés A, B et C.



▲ **Figure 9.3** **Comparaison d'estimateurs**

La vraie valeur de θ est fixée à 2 (barre verticale). Intuitivement, les réalisations de l'estimateur A ont de grandes chances d'être assez éloignées de la vraie valeur de θ. En moyenne ses réalisations sont centrées sur 4, valeur qui correspond à son espérance. On dit que l'estimateur A est **biaisé** : son espérance est différente de la vraie valeur du paramètre à estimer. En revanche, les deux estimateurs B et C ne sont pas biaisés. Si l'on effectue plusieurs tirages (réalisations) dans ces deux distributions, on obtiendra *en moyenne* des estimations (réalisations) égales à 2. Pour autant, on préférera sans conteste l'estimateur C à l'estimateur B. Sa distribution est beaucoup plus concentrée autour de la vraie valeur de θ que celle de l'estimateur B.

Quelle est l'implication de ce résultat ? Si l'on effectue des tirages dans la distribution de l'estimateur C, on obtiendra par exemple des valeurs du type 2,2, 1,9, 2,10, 1,8, etc., valeurs assez proches de la vraie valeur de θ relativement à celles que l'on pourrait obtenir avec l'estimateur B (par exemple 1,2, 3,2, 2,8, 0,8, etc.). L'estimateur C est donc plus **précis** que l'estimateur B, parce que sa **variance** est plus faible. Sur la base de cette comparaison théorique des **distributions d'échantillonnage**, on préférera donc l'estimateur C aux estimateurs A et B. Ainsi, c'est cet estimateur que l'on appliquera sur un échantillon pour obtenir, au final, l'estimation du paramètre θ.

L'étude des propriétés d'un estimateur est basée sur l'étude des caractéristiques de sa **distribution**. Toutefois, nous devons distinguer deux cas suivant la taille d'échantillon n :

- l'étude de la distribution et des propriétés à **distance finie** pour n fixe ;
- l'étude de la distribution et des propriétés **asymptotiques** pour $n \to \infty$.

3 Propriétés à distance finie

Les propriétés à distance finie d'un estimateur correspondent aux propriétés de sa distribution à distance finie obtenue pour un échantillon de taille n finie.

3.1 Distribution à distance finie

Définition 9.9

La **distribution à distance finie** (ou **distribution exacte**) d'un estimateur $\widehat{\theta}$ correspond à la distribution valable pour toute valeur de la taille de l'échantillon $n \in \mathbb{N}$:

$$\widehat{\theta} \sim \text{loi exacte}(n) \quad \forall n \in \mathbb{N}$$

La distribution exacte est nécessairement paramétrée par la taille d'échantillon n. Considérons un exemple de distribution à distance finie.

Exemple

On dispose d'un n-échantillon $(X_1,\ldots,X_n)$ où les variables aléatoires X_i sont N.i.d.$\left(\mu,\sigma^2\right)$. Cet acronyme signifie que ces variables sont *normalement* et indépendamment distribuées. La *distribution à distance finie* de l'estimateur $\widehat{\mu} = \overline{X}_n$ (moyenne empirique) de l'espérance μ est la suivante :

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \quad \forall n \in \mathbb{N} \tag{9.17}$$

En effet, nous savons que la somme de variables aléatoires normales indépendantes suit une loi normale. Par ailleurs, nous savons que :

$$\mathbb{E}(\widehat{\mu}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{n \times \mu}{n} = \mu \tag{9.18}$$

$$\begin{aligned}\mathbb{V}(\widehat{\mu}) &= \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\\ &= \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathbb{V}(X_i) + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1}\mathbb{C}ov\left(X_i,X_j\right)\right)\\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i) = \frac{n \times \sigma^2}{n^2} = \frac{\sigma^2}{n}\end{aligned} \tag{9.19}$$

puisque (1) les variables X_i sont identiquement distribuées avec $\mathbb{E}(X_i) = \mu$ et $\mathbb{V}(X_i) = \sigma^2$ pour $i = 1,\ldots,n$, (2) les variables X_i sont indépendantes, impliquant $\mathbb{C}ov\left(X_i,X_j\right) = 0$ pour $i \neq j$. La distribution de $\widehat{\mu}$ pour toute taille d'échantillon $n \in \mathbb{N}$ est totalement caractérisée par μ et σ^2, paramètres qui peuvent être estimés. Par exemple, si $n = 3$, alors $\widehat{\mu} \sim \mathcal{N}\left(\mu,\sigma^2/3\right)$, si $n = 10$, alors $\widehat{\mu} \sim \mathcal{N}\left(\mu,\sigma^2/10\right)$, etc.

Il est souvent très compliqué de déterminer la distribution exacte d'un estimateur. Parfois, on peut seulement déterminer la distribution exacte d'une **variable transformée** de l'estimateur $\widehat{\theta}$. Considérons l'exemple suivant.

Exemple

On dispose d'un n-échantillon de variables aléatoires *normales* $(X_1,\ldots,X_n)$ où les variables X_i sont N.i.d.$\left(\mu,\sigma^2\right)$. La variance empirique

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 \tag{9.20}$$

est un estimateur de la variance σ^2, avec $\overline{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. On admet que, sous ces hypothèses, la variable aléatoire transformée $(n-1)\, S_n^2/\sigma^2$ a une *distribution exacte* du khi-deux à $n-1$ degrés de liberté :

$$\frac{(n-1)}{\sigma^2} S_n^2 \sim \chi^2(n-1) \quad \forall n \in \mathbb{N} \tag{9.21}$$

Dans ce cas on ne connaît pas la distribution exacte de l'estimateur S_n^2, mais l'on connaît celle de la variable transformée $(n-1)\, S_n^2/\sigma^2$.

Mais sauf dans des cas particuliers (échantillon normal), il est généralement **impossible de déterminer la distribution exacte** de l'estimateur $\widehat{\theta}$ ou d'une transformée de cet estimateur. Deux raisons expliquent cela :

Première raison. Dans certains cas, on connaît la distribution des variables $X_1,...,X_n$ de l'échantillon, par exemple $X_i \sim \mathcal{N}(\mu,\sigma^2)$ ou $X_i \sim \chi^2(v)$, etc. Mais la fonction $g(.)$ qui définit l'estimateur est trop compliquée pour permettre la dérivation de la distribution exacte de $\widehat{\theta}$.

$$\widehat{\theta} = g(X_1,...,X_n) \sim ? \quad \forall n \in \mathbb{N} \tag{9.22}$$

Ainsi dans l'exemple précédent, la distribution de la variance empirique corrigée S_n^2 est inconnue, y compris dans le cas d'un échantillon normal.

Deuxième raison. Dans la plupart des cas, on ne connaît pas la distribution des variables $X_1,...,X_n$ de l'échantillon. Si ces variables sont indépendantes et identiquement distribuées (i.i.d), tout ce que l'on sait c'est qu'elles ont la même distribution, mais cette dernière est *a priori* inconnue. Dès lors, l'estimateur $\widehat{\theta}$, défini comme une fonction de ces variables aléatoires de distribution inconnue, a lui-même une distribution (à distance finie) inconnue.

$$\widehat{\theta} = g(X_1,...,X_n) \sim ? \quad \forall n \in \mathbb{N} \tag{9.23}$$

À ce stade de l'exposé, la question qui se pose est de savoir comment évaluer la **performance** d'un estimateur. Pour cela, nous allons nous intéresser aux **moments** de la distribution de $\widehat{\theta}$. Ces moments permettent de caractériser certaines **propriétés à distance finie** de l'estimateur : son biais (espérance), sa précision (variance), etc.

Il est souvent possible de déterminer ces moments même si la distribution exacte de l'estimateur est inconnue. En imposant certaines hypothèses, on peut par exemple calculer $\mathbb{E}(\widehat{\theta})$, $\mathbb{V}(\widehat{\theta})$, etc., sans connaître la forme de la densité de l'estimateur $\widehat{\theta}$.

3.2 Le biais d'un estimateur

Le premier moment de la distribution de $\widehat{\theta}$, *i.e.* l'espérance, détermine son biais[7].

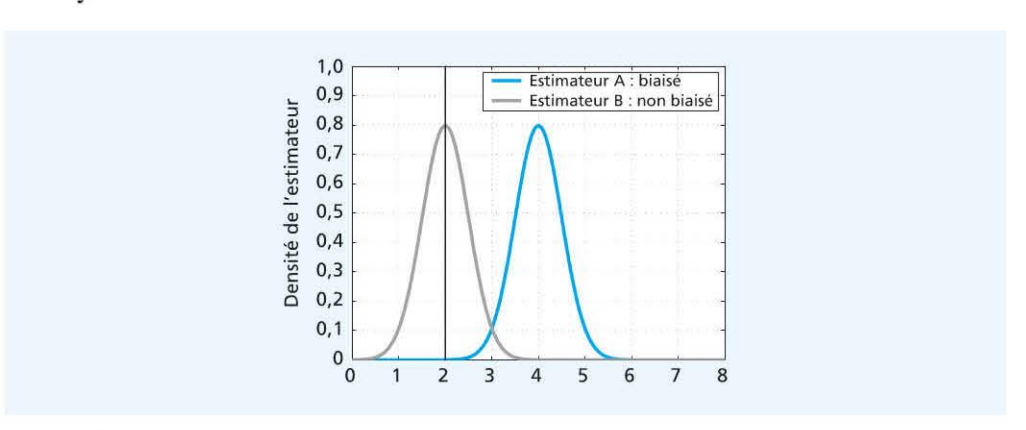
Définition 9.10

Un estimateur $\widehat{\theta}$ d'un paramètre θ est **non biasé** si l'espérance de sa distribution est égale à θ :

$$\mathbb{E}(\widehat{\theta}) = \theta \tag{9.24}$$

7 On peut définir le biais d'un estimateur par la quantité $\text{Biais}(\widehat{\theta}) = \mathbb{E}(\widehat{\theta} - \theta)$.

Considérons l'illustration de la figure 9.4 où sont représentées les fonctions de densité de deux estimateurs A et B. On observe que l'estimateur A est biaisé car l'espérance de sa distribution d'échantillonnage, $\mathbb{E}(\widehat{\theta_A}) = 4$, est différente de la vraie valeur ($\theta = 2$) du paramètre. Si l'on effectue plusieurs tirages dans cette distribution, la moyenne des réalisations de $\widehat{\theta_A}$ (estimations du paramètre θ) sera différente de la vraie valeur de θ. En revanche l'estimateur B est non biaisé : l'espérance de sa distribution coïncide avec la vraie valeur de θ. Ainsi, les réalisations de $\widehat{\theta_B}$ (estimations du paramètre θ) seront en moyenne centrées sur 2.



▲ **Figure 9.4 Estimateurs biaisé et non biaisé**

Exemple

Soit $(Y_1,\ldots,Y_n)$ un n-échantillon de variables aléatoires discrètes i.i.d. telles que Y_i admette une distribution de Bernoulli avec une probabilité de succès $p \in [0,1]$. La moyenne empirique est un estimateur sans biais de p :

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{9.25}$$

En effet, puisque les variables Y_i sont i.i.d. avec $\mathbb{E}(Y_i) = p$, on a :

$$\mathbb{E}(\widehat{p}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(Y_i) = \frac{p \times n}{n} = p \tag{9.26}$$

Exemple

Soit $(Y_1,\ldots,Y_n)$ un n-échantillon de variables aléatoires continues i.i.d. admettant une distribution uniforme $\mathcal{U}_{[0,\theta]}$. Un estimateur sans biais de θ est :

$$\widehat{\theta} = \frac{2}{n}\sum_{i=1}^{n} Y_i \tag{9.27}$$

En effet, puisque les variables Y_i sont i.i.d. avec $\mathbb{E}(Y_i) = (\theta + 0)/2 = \theta/2$, on a :

$$\mathbb{E}\left(\widehat{\theta}\right) = \mathbb{E}\left(\frac{2}{n}\sum_{i=1}^{n} Y_i\right) = \frac{2}{n}\sum_{i=1}^{n} \mathbb{E}(Y_i) = \frac{2}{n} \times \frac{n \times \theta}{2} = \theta \tag{9.28}$$

Dans les deux exemples précédents, nous avons pu montrer que l'estimateur était sans biais, sans connaître pour autant sa distribution exacte. Dans l'exemple suivant, nous allons au contraire utiliser cette distribution, ou plus précisément la distribution d'une variable transformée de l'estimateur.

Exemple

Soit $(Y_1,\ldots,Y_n)$ un n-échantillon de variables N.i.d. Montrons que la variance empirique non corrigée :

$$\widetilde{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \overline{Y}_n\right)^2 \tag{9.29}$$

est un estimateur *biaisé* de la variance σ^2. Pour cela, introduisons tout d'abord la variance empirique corrigée S_n^2 :

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \overline{Y}_n\right)^2 \tag{9.30}$$

Nous connaissons la distribution exacte d'une forme transformée de S_n^2 puisque :

$$\left(\frac{n-1}{\sigma^2}\right)S_n^2 \sim \chi^2(n-1) \quad \forall n \in \mathbb{N} \tag{9.31}$$

Exprimons maintenant la variance empirique corrigée S_n^2 en fonction de $\widetilde{S}_n^2$:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \overline{Y}_n\right)^2 = \left(\frac{n}{n-1}\right)\widetilde{S}_n^2 \tag{9.32}$$

Par conséquent il vient :

$$\left(\frac{n-1}{\sigma^2}\right)S_n^2 = \left(\frac{n-1}{\sigma^2}\right)\left(\frac{n}{n-1}\right)\widetilde{S}_n^2 = \frac{n}{\sigma^2}\widetilde{S}_n^2 \tag{9.33}$$

On en déduit que :

$$\frac{n}{\sigma^2}\widetilde{S}_n^2 \sim \chi^2(n-1) \tag{9.34}$$

On sait que si $X \sim \chi^2(v)$ alors $\mathbb{E}(X) = v$. Dès lors :

$$\mathbb{E}\left(\frac{n}{\sigma^2}\widetilde{S}_n^2\right) = n-1 \Longleftrightarrow \frac{n}{\sigma^2}\mathbb{E}\left(\widetilde{S}_n^2\right) = n-1 \tag{9.35}$$

ou de façon équivalente :

$$\mathbb{E}\left(\widetilde{S}_n^2\right) = \left(\frac{n-1}{n}\right)\sigma^2 \neq \sigma^2 \tag{9.36}$$

Ainsi, la variance empirique non corrigée $\widetilde{S}_n^2$ est un estimateur *biaisé* de σ^2. On remarque que lorsque $n \to \infty$, $\mathbb{E}(\widetilde{S}_n^2) = \sigma^2$. On dit que cet estimateur est *asymptotiquement non biaisé*.

FOCUS

La variance empirique corrigée

On distingue les **variances empiriques** corrigée et non corrigée. Cette **correction** est parfois appelée correction de petit échantillon. Soit un n-échantillon $(X_1,\ldots,X_n)$ de variables aléatoires i.i.d. telles que $\mathbb{E}(X_i) = \mu$ et $\mathbb{V}(X_i) = \sigma^2$, les variances empiriques corrigée et non corrigée sont respectivement définies par :

$$S_n^2 = \underbrace{\frac{1}{n-1}}_{\text{correction}}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2, \quad \widetilde{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 \tag{9.37}$$

La variance empirique non corrigée $\widetilde{S}_n^2$ est un estimateur biaisé de σ^2, tandis que la variance empirique corrigée S_n^2 est un estimateur sans biais :

$$\mathbb{E}(\widetilde{S}_n^2) \neq \sigma^2, \qquad \mathbb{E}\left(S_n^2\right) = \sigma^2 \tag{9.38}$$

L'intuition de cette correction est la suivante. Supposons que les variables X_i soient normalement distribuées avec $\mu = 0$, *i.e.* $X_i \sim \mathcal{N}\left(0,\sigma^2\right)$. Dans ce cas, la variable X_i/σ suit une loi normale centrée réduite et la variable X_i^2/σ^2 suit une loi du khi-deux à un degré de liberté. Par conséquent, la somme de ces variables *indépendantes* pour $i = 1,\ldots,n$ suit une loi du khi-deux à n degrés de liberté.

$$\frac{1}{\sigma^2}\sum_{i=1}^{n} X_i^2 \sim \chi^2(n) \tag{9.39}$$

Considérons la « pseudo »-variance S^2 : $S^2 = n^{-1}\sum_{i=1}^{n} X_i^2$. Par définition :

$$\frac{1}{\sigma^2}\sum_{i=1}^{n} X_i^2 = \frac{n}{\sigma^2} S^2 \sim \chi^2(n) \tag{9.40}$$

Étant données les propriétés de la loi du khi-deux, on en déduit que :

$$\mathbb{E}\left(\frac{n}{\sigma^2} S^2\right) = n \Longleftrightarrow \mathbb{E}\left(S^2\right) = \sigma^2 \tag{9.41}$$

Sous l'hypothèse $\mu = 0$, la variable S^2 est un estimateur sans biais de σ^2. Mais cette variable ne correspond pas à la définition de la variance empirique $\widetilde{S}_n^2$ (équation 9.37). En effet, la variance empirique dépend de la somme $\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$ et non de $\sum_{i=1}^{n} X_i^2$. Le problème c'est que les variables $\left(X_i - \overline{X}_n\right)^2$ pour $i = 1,\ldots,n$ ne sont *pas indépendantes* en raison de la présence du terme $\overline{X}_n$. On peut montrer que seules $n-1$ variables sont indépendantes, d'où :

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 = \frac{\sigma^2}{n}\widetilde{S}_n^2 \sim \chi^2(n-1) \tag{9.42}$$

C'est pourquoi $\widetilde{S}_n^2$ est un estimateur biaisé (en petit échantillon) de σ^2 :

$$\mathbb{E}\left(\widetilde{S}_n^2\right) = \left(\frac{n-1}{n}\right)\sigma^2 \neq \sigma^2 \tag{9.43}$$

L'absence de biais n'est toutefois pas un critère suffisant pour discriminer des estimateurs alternatifs. Pour un même problème, on peut facilement trouver plusieurs estimateurs sans biais.

Exemple

Soit un n-échantillon $(Y_1,\ldots,Y_n)$ de variables aléatoires i.i.d. telles que $\mathbb{E}(Y_i) = \mu$. On considère deux estimateurs $\widehat{\mu}_1$ et $\widehat{\mu}_2$ de l'espérance μ. Le premier estimateur correspond à la moyenne empirique et le deuxième estimateur n'est rien d'autre que la première variable de l'échantillon :

$$\widehat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} Y_i \qquad \widehat{\mu}_2 = Y_1 \tag{9.44}$$

Ces deux estimateurs sont des estimateurs sans biais de μ. En effet :

$$\mathbb{E}(\widehat{\mu}_1) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Y_i) = \frac{n \times \mu}{n} = \mu \tag{9.45}$$

$$\mathbb{E}(\widehat{\mu}_2) = \mathbb{E}(Y_1) = \mu \tag{9.46}$$

puisque les variables Y_i sont i.i.d. avec $\mathbb{E}(Y_i) = \mu$.

3.3 Précision et efficacité d'un estimateur

Comment comparer deux **estimateurs non biaisés** ? Cette comparaison se fait sur la base de leur **variance**.

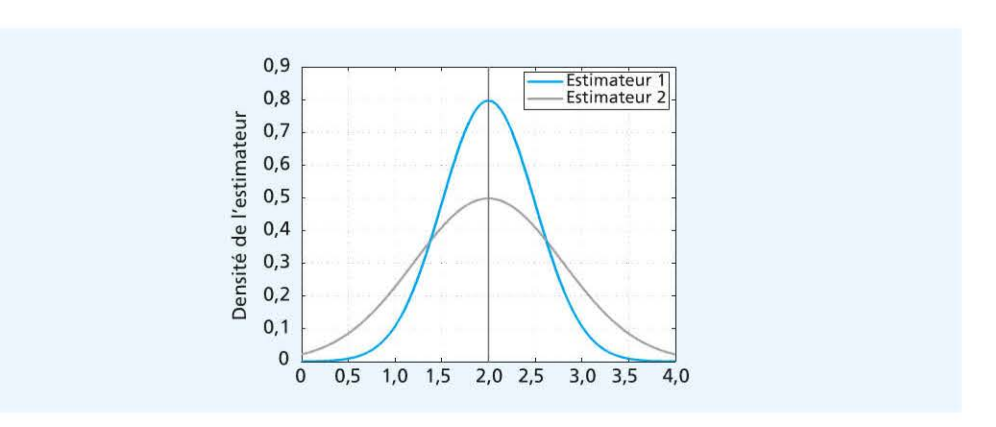
Propriété

Comparaison d'estimateurs sans biais

Soient deux estimateurs sans biais $\widehat{\theta_1}$ et $\widehat{\theta_2}$. L'estimateur $\widehat{\theta_1}$ domine l'estimateur $\widehat{\theta_2}$, *i.e.* $\widehat{\theta_1} \geq \widehat{\theta_2}$, si :

$$\mathbb{V}\left(\widehat{\theta_1}\right) \leq \mathbb{V}\left(\widehat{\theta_2}\right) \tag{9.47}$$

Comme l'illustre la figure 9.5, l'idée est que plus la variance d'un estimateur sans biais est faible, plus sa densité est concentrée autour de la vraie valeur du paramètre, plus les estimations ont de fortes chances d'être proches de cette valeur.



▲ **Figure 9.5** Comparaison d'estimateurs sans biais

Exemple

Soit un n-échantillon $(Y_1,\ldots,Y_n)$ i.i.d. tel que $\mathbb{E}(Y_i) = \mu$ et $\mathbb{V}(Y_i) = \sigma^2$. Comparons les deux estimateurs $\widehat{\mu}_1 = n^{-1}\sum_{i=1}^{n} Y_i$ et $\widehat{\mu}_2 = Y_1$ de l'espérance μ. Tout d'abord, nous savons que ces deux estimateurs sont « sans biais ». Par ailleurs :

$$\mathbb{V}(\widehat{\mu}_1) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(Y_i) = \frac{\sigma^2}{n} \tag{9.48}$$

$$\mathbb{V}(\widehat{\mu}_2) = \mathbb{V}(Y_1) = \sigma^2 \tag{9.49}$$

On obtient $\mathbb{V}(\widehat{\mu}_1) \leq \mathbb{V}(\widehat{\mu}_2)$ dès lors que la taille d'échantillon n est supérieure ou égale à un. L'estimateur $\widehat{\mu}_1$ est préféré à $\widehat{\mu}_2$.

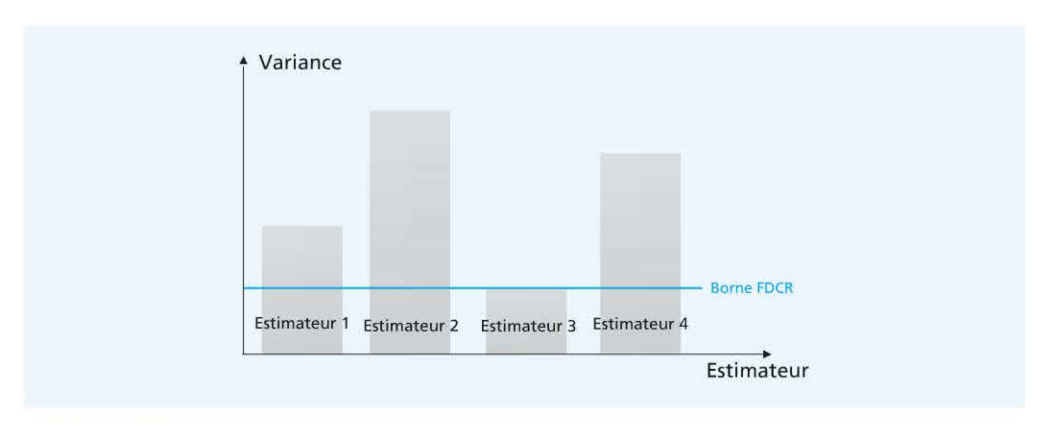
Remarque : Seuls des estimateurs non biaisés peuvent être comparés sur la base de leur variance.

Ainsi, nous savons comparer deux estimateurs non baisés. Mais existe-t-il un estimateur sans biais qui soit plus efficace que tous les autres ? C'est la notion d'estimateur **optimal**.

Définition 9.11

Un estimateur **optimal** au sens du critère de la variance (ou de l'erreur quadratique) est l'estimateur sans biais qui possède la variance la plus faible parmi tous les estimateurs sans biais.

Il est souvent difficile, voire impossible, de montrer qu'un estimateur est optimal. Une alternative consiste à montrer que la variance d'un estimateur atteint une certaine **borne** en deçà de laquelle les variances des estimateurs sans biais ne peuvent pas descendre, comme l'illustre la figure 9.6. C'est le concept de borne de **Cramer-Rao** ou de **borne FDCR (Frechet - Darnois - Cramer - Rao)**.



▲ **Figure 9.6 Illustration du concept de borne FDCR**

Il est important de noter que la borne FDCR ne peut être établie que sous un certain nombre d'**hypothèses** (► chapitre 10), c'est pourquoi le concept d'efficacité au sens FDCR est plus restrictif que le concept d'optimalité, même si l'idée est similaire.

Propriété

Borne FDCR

Soit $(X_1,\ldots,X_n)$ un échantillon i.i.d. où X_i admet une fonction de densité (ou de masse) $f_X(\theta;x)$ dépendant d'un paramètre θ. Soit $\widehat{\theta}$ un estimateur sans biais de θ, *i.e.*, $\mathbb{E}(\widehat{\theta}) = \theta$. Si la fonction $f_X(\theta;x)$ est régulière alors :

$$\mathbb{V}\left(\widehat{\theta}\right) \geq \mathrm{I}_n^{-1}(\theta_0) = \textbf{borne FDCR ou borne de Cramer-Rao} \tag{9.50}$$

où $\mathrm{I}_n(\theta_0)$ correspond à la quantité d'information de Fisher associée à l'échantillon et évaluée en θ_0, vraie valeur du paramètre θ.

Dans le chapitre 10, consacré au maximum de vraisemblance, nous reviendrons sur cette définition après avoir introduit le concept de quantité d'information de Fisher et la notion de fonction de densité régulière. Mais ce qu'il faut retenir dès à présent, c'est que si la variance d'un estimateur atteint cette borne, on dit que cet estimateur est **efficace** (au sens de la borne FDCR), puisqu'il domine tous les autres estimateurs sans biais, en termes de variance.

Définition 9.12

Un estimateur est **efficace** au sens de la borne FDCR (Frechet - Darnois - Cramer - Rao) ou de la borne Cramer-Rao, si :

$$\mathbb{V}\left(\widehat{\theta}\right) = \mathrm{I}_n^{-1}(\theta_0) \tag{9.51}$$

où $\mathrm{I}_n(\theta_0)$ correspond à la quantité d'information de Fisher associée à l'échantillon et évaluée pour la vraie valeur θ_0 du paramètre θ.

On dit aussi qu'un estimateur efficace est BUE (*Best Unbiased Estimator*). Cela traduit le fait que c'est le meilleur des estimateurs sans biais en termes de variance. Lorsqu'il est impossible de déterminer la borne FDCR (▶ chapitre 10), on a parfois recours au concept d'estimateur BLUE (*Best Linear Unbiased Estimator*). On caractérise alors le meilleur des estimateurs sans biais parmi la classe des estimateurs linéaires, un estimateur linéaire étant défini comme une somme pondérée des variables de l'échantillon.

3.4 Extension au cas multivarié

Jusqu'ici, nous avons considéré le cas où le paramètre à estimer, θ, était un scalaire. Nous allons à présent étendre les définitions précédentes au cas où θ est un **vecteur de k paramètres** :

$$\theta = (\theta_1,\ldots,\theta_k)^\top \tag{9.52}$$

Le symbole $^\top$ correspond à la transposée. Considérons quelques exemples.

Exemple

Soit $(X_1,\ldots,X_n)$ un échantillon de variables N.i.d. avec $\mathbb{E}(X_i) = \mu$ et $\mathbb{V}(X_i) = \sigma^2$. Les deux paramètres μ et σ^2 sont inconnus et l'on cherche à les estimer. On pose $\theta = \left(\mu,\sigma^2\right)^\top$ et $k = 2$.

On considère un modèle de régression (▶ chapitre 2) :

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\mu} \tag{9.53}$$

où $\boldsymbol{y} = (y_1,\ldots,y_n)^\top \in \mathbb{R}^n$ est un vecteur de variables aléatoires endogènes, $\boldsymbol{X}$ est une matrice de dimension $n \times k$ de régresseurs non stochastiques, et $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_n)^\top \in \mathbb{R}^n$ est un vecteur de termes d'erreur aléatoires tel que $\mathbb{E}(\boldsymbol{\mu}) = 0_{n\times 1}$ et $\mathbb{V}(\boldsymbol{\mu}) = \sigma^2 \boldsymbol{I}_n$ où I_n est la matrice identité de dimension $n \times n$. Le vecteur de paramètres $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_k)^\top$ est inconnu. On pose donc $\theta = \boldsymbol{\beta}$.

On considère un estimateur $\widehat{\theta}$, défini par un vecteur de dimension $k \times 1$, tel que :

$$\underset{(k\times 1)}{\widehat{\theta}} = \begin{pmatrix} \widehat{\theta}_1 \\ \ldots \\ \widehat{\theta}_k \end{pmatrix} \tag{9.54}$$

Son *espérance* $\mathbb{E}\left(\widehat{\theta}\right)$ est un vecteur de $k \times 1$ valeurs :

$$\underset{(k\times 1)}{\mathbb{E}\left(\widehat{\theta}\right)} = \begin{pmatrix} \mathbb{E}\left(\widehat{\theta}_1\right) \\ \ldots \\ \mathbb{E}\left(\widehat{\theta}_k\right) \end{pmatrix} \tag{9.55}$$

Ainsi, cet estimateur est *non biaisé* si et seulement si :

$$\underset{(k\times 1)}{\mathbb{E}\left(\widehat{\theta}\right)} = \underset{(k\times 1)}{\theta} \Longleftrightarrow \begin{pmatrix} \mathbb{E}\left(\widehat{\theta_1}\right) \\ \ldots \\ \mathbb{E}\left(\widehat{\theta_k}\right) \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \ldots \\ \theta_k \end{pmatrix} \tag{9.56}$$

La **matrice de variance-covariance** de l'estimateur $\widehat{\theta}$ est de dimension $k \times k$:

$$\underset{(k\times k)}{\mathbb{V}\left(\widehat{\theta}\right)} = \begin{pmatrix} \mathbb{V}\left(\widehat{\theta_1}\right) & \mathbb{C}ov\left(\widehat{\theta_1},\widehat{\theta_2}\right) & \ldots & \mathbb{C}ov\left(\widehat{\theta_1},\widehat{\theta_k}\right) \\ \mathbb{C}ov\left(\widehat{\theta_2},\widehat{\theta_1}\right) & \mathbb{V}\left(\widehat{\theta_2}\right) & \ldots & \mathbb{C}ov\left(\widehat{\theta_2},\widehat{\theta_k}\right) \\ \ldots & \ldots & \ldots & \ldots \\ \mathbb{C}ov\left(\widehat{\theta_k},\widehat{\theta_1}\right) & \mathbb{C}ov\left(\widehat{\theta_k},\widehat{\theta_2}\right) & \ldots & \mathbb{V}\left(\widehat{\theta_k}\right) \end{pmatrix} \tag{9.57}$$

Remarque : La matrice de variance-covariance de l'estimateur $\widehat{\theta}$, comme toute matrice de variance-covariance est symétrique, inversible et définie-positive (c'est-à-dire que ses valeurs propres sont toutes positives).

Dans le cas multivarié, la notion de distribution d'échantillonnage d'un estimateur doit être précisée. La distribution exacte de l'estimateur $\widehat{\theta}$, si elle existe, est une **distribution multivariée** ou **distribution jointe** (▶ chapitre 6). Par exemple, on peut avoir $\widehat{\theta} \sim \mathcal{N}\left(\theta, n^{-1}\Sigma\right)$ où Σ est une matrice de variance-covariance de dimension $k \times k$. On peut alors déterminer les **lois marginales** des estimateurs individuels $\widehat{\theta_1},..,\widehat{\theta_k}$ ainsi que les **distributions conditionnelles**. Par exemple, on peut établir la distribution conditionnelle de $\widehat{\theta_1}$ sachant que l'estimateur $\widehat{\theta_2}$ est égal à une valeur c.

Revenons sur la comparaison d'estimateurs non biaisés. Soient deux estimateurs $\widehat{\theta_1}$ et $\widehat{\theta_2}$. L'estimateur $\widehat{\theta_1}$ est *préféré* à l'estimateur $\widehat{\theta_2}$ si et seulement si :

$$\underbrace{\mathbb{V}\left(\widehat{\theta_2}\right) - \mathbb{V}\left(\widehat{\theta_1}\right)}_{(k\times k)} \text{ est une matrice semi-définie positive} \tag{9.58}$$

Cette expression se réduit à $\mathbb{V}(\widehat{\theta_2}) - \mathbb{V}(\widehat{\theta_1}) \geq 0$ dans le cas univarié. De la même façon, un estimateur est *efficace* au sens de la borne FDCR si et seulement si :

$$\underset{(k\times k)}{\mathbb{V}\left(\widehat{\theta}\right)} = \underset{(k\times k)}{\mathrm{I}_n^{-1}(\theta_0)} \tag{9.59}$$

où $\mathrm{I}_n(\theta_0)$ désigne la matrice d'information de Fisher associée à l'échantillon et évaluée au point θ_0, vraie valeur du paramètre θ.

4 Propriétés asymptotiques

La question qui se pose ici est de savoir comment se comporte l'estimateur $\widehat{\theta}$ lorsque la taille d'échantillon n tend vers l'**infini**. Pourquoi étudier le **comportement asymptotique** de $\widehat{\theta}$? Dans la plupart des problèmes, il est impossible de dériver la distribution exacte de $\widehat{\theta}$, c'est-à-dire sa distribution valable pour toute valeur de n. C'est en particulier vrai lorsque l'on ne connaît pas la distribution des variables de l'échantillon (par exemple, lorsque ces variables sont supposées i.i.d. de distribution commune

inconnue) ou que la forme fonctionnelle de l'estimateur, c'est-à-dire la fonction $g(X_1,...,X_n)$, est trop compliquée. Dans ce contexte, on cherche à caractériser le comportement de la variable aléatoire $\widehat{\theta}$ dans le cas hypothétique d'un échantillon de taille infinie à l'aide des différentes notions de convergence (en probabilité, presque sûre, en moyenne quadratique ou en loi, ► chapitre 8). On étudie généralement deux dimensions :

- la **convergence** de l'estimateur $\widehat{\theta}$;
- la **distribution asymptotique** de $\widehat{\theta}$, généralement établie à partir du théorème central limite (► chapitre 8).

4.1 Estimateur convergent

Soit un estimateur $\widehat{\theta_n} = g(X_1,...,X_n)$ d'un paramètre (ou d'un vecteur de paramètres) θ associé à un n-échantillon $(X_1,...,X_n)$. Afin de bien mettre en évidence la dépendance de l'estimateur à la taille de l'échantillon, nous l'indicerons par n et nous noterons θ_0 la vraie valeur du paramètre θ.

Définition 9.13

Un estimateur $\widehat{\theta_n}$ est **convergent au sens fort** s'il converge presque sûrement vers la vraie valeur du paramètre :

$$\widehat{\theta_n} \overset{a.s.}{\to} \theta_0 \tag{9.60}$$

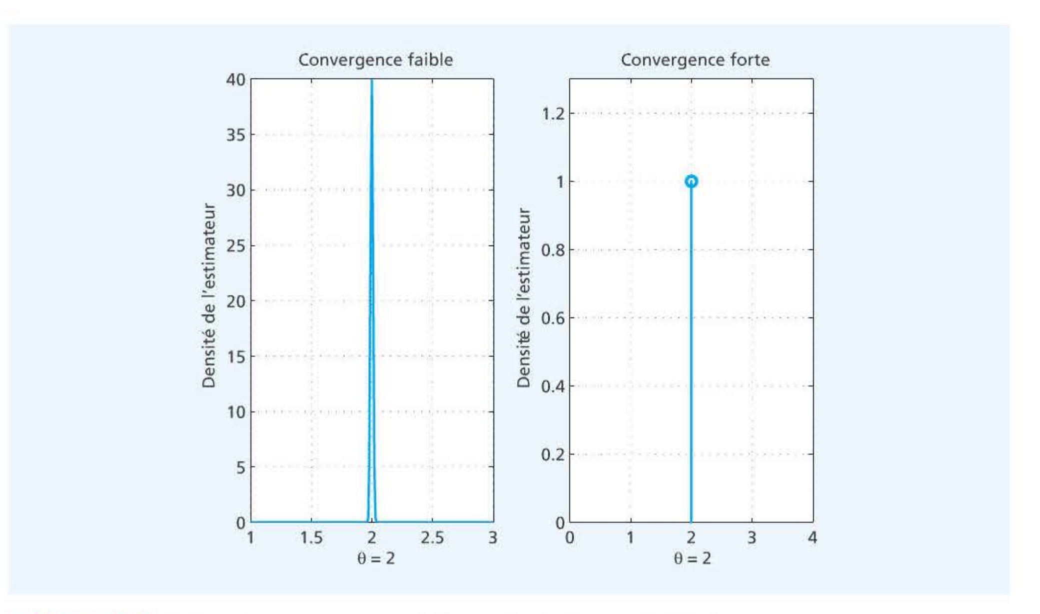
Définition 9.14

Un estimateur $\widehat{\theta_n}$ est **convergent au sens faible** s'il converge en probabilité vers la vraie valeur du paramètre :

$$\widehat{\theta_n} \overset{p}{\to} \theta_0 \tag{9.61}$$

Remarque : Lorsqu'un estimateur est qualifié de convergent sans plus de précision (*consistent* en anglais), cela signifie qu'il est convergent au sens faible.

La **convergence** est une des propriétés les plus importantes pour un estimateur. Elle signifie que si l'on applique l'estimateur à un très grand échantillon, les estimations (*i.e.* les réalisations de $\widehat{\theta_n}$) seront extrêmement **concentrées** autour de la vraie valeur du paramètre. Comme l'illustre la partie gauche de la figure 9.7, dans le cas d'une convergence faible, lorsque n tend vers l'infini il y a une très forte probabilité d'obtenir des estimations $\widehat{\theta_n}(x)$ très proches de la vraie valeur de θ. Si cette valeur est égale à 2, on obtiendra par exemple des estimations du type 2,0001, 1,9999, 2,0000, etc. Dans le cas de la convergence forte (graphique de droite de la figure 9.7), les choses sont encore plus simples. Lorsque n tend vers l'infini, la distribution de l'estimateur $\widehat{\theta_n}$ est dégénérée en une masse ponctuelle : l'estimateur n'est plus une variable aléatoire et devient une constante égale à la vraie valeur du paramètre, $\theta = 2$. On comprend dès lors l'intérêt de démontrer la convergence d'un estimateur, au minimum au sens faible.



▲ **Figure 9.7 Estimateur convergent (sens fort et sens faible)**

Propriété

Convergence au sens faible

Soit un estimateur $\widehat{\theta_n}$ d'un paramètre (ou d'un vecteur de paramètres) θ tel que :

$$\lim_{n\to\infty} \mathbb{E}\left(\widehat{\theta_n}\right) = \theta_0 \qquad \lim_{n\to\infty} \mathbb{V}\left(\widehat{\theta_n}\right) = 0 \tag{9.62}$$

où θ_0 est la vraie valeur du paramètre, alors cet estimateur est **convergent** au sens faible (► chapitre 8) :

$$\widehat{\theta_n} \overset{p}{\to} \theta_0 \tag{9.63}$$

Exemple

Soit $(Y_1,\ldots,Y_n)$ un n-échantillon de variables aléatoires i.i.d. telles que $\mathbb{E}(Y_i) = \mu$ et $\mathbb{V}(Y_i) = \sigma^2$, où le paramètre μ est inconnu. L'estimateur $\widehat{\mu_n}$, défini par :

$$\widehat{\mu_n} = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{9.64}$$

est un estimateur convergent de μ. En effet, puisque les variables $Y_1, Y_2, \ldots, Y_n$ sont i.i.d., nous avons :

$$\mathbb{E}(\widehat{\mu_n}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(Y_i) = \mu \tag{9.65}$$

$$\lim_{n\to\infty} \mathbb{V}(\widehat{\mu_n}) = \lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{V}(Y_i) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0 \tag{9.66}$$

L'estimateur $\widehat{\mu_n}$ est donc convergent au sens faible :

$$\widehat{\mu_n} \overset{p}{\to} \mu$$

Une autre façon de démontrer la convergence en probabilité consiste à utiliser la **loi faible des grands nombres** (théorème de Khintchine). Dans le cadre d'un échantillon i.i.d., nous savons que la moyenne empirique $\overline{X}_n$ des variables de l'échantillon

converge vers l'espérance. Il suffit alors d'exprimer $\widehat{\theta}_n$ comme une fonction de cette moyenne empirique, *i.e.* sous la forme $\widehat{\theta}_n = h(\overline{X}_n)$. En utilisant le théorème de Slutsky (▶ chapitre 8), on en déduit la convergence en probabilité de $\widehat{\theta}_n$ et son éventuel caractère convergent ou non.

4.2 Distribution asymptotique

Dans la plupart des problèmes d'estimation mis en œuvre dans la pratique, on cherche à déterminer la **distribution asymptotique** de l'estimateur. La loi asymptotique permet notamment d'estimer l'écart-type (ou *standard error* en anglais) associé à un estimateur dont on ne connaît pas nécessairement la distribution exacte (▶ En pratique : Estimation et logiciel d'économétrie).

Définition 9.15

Au sens strict, la **distribution asympotique** d'un estimateur $\widehat{\theta}_n$ correspond à sa distribution valable uniquement pour une taille d'échantillon n très importante mais finie.

Pourquoi s'intéresser à la distribution de l'estimateur dans le cas particulier où la taille de l'échantillon n est très importante mais finie ? D'un coté, nous savons qu'il est généralement impossible (sauf sous des hypothèses fortes portant sur la normalité de l'échantillon) de connaître la distribution exacte (à distance finie) de l'estimateur, valable quelle que soit la dimension finie n de l'échantillon :

$$\widehat{\theta}_n \sim \text{loi exacte ??} \quad \forall n \in \mathbb{N} \tag{9.67}$$

D'un autre coté, lorsque la taille de l'échantillon n tend vers l'infini, si l'estimateur $\widehat{\theta}_n$ est convergent (au sens strict ou au sens faible), alors sa distribution tend vers une distribution dégénérée. Par exemple, dans le cas d'un estimateur convergent au sens fort, $\widehat{\theta}_n \overset{a.s.}{\to} \theta_0$, la densité $f_{\widehat{\theta}_n}(x)$ de $\widehat{\theta}_n$ tend vers une masse ponctuelle :

$$\lim_{n\to\infty} f_{\widehat{\theta}_n}(x) = f(x) = \begin{cases} 1 & \text{si } x = \theta_0 \\ 0 & 0 \text{ sinon} \end{cases} \tag{9.68}$$

où θ_0 désigne la vraie valeur du paramètre. Dit autrement, lorsque n tend vers l'infini, l'estimateur $\widehat{\theta}_n$ converge vers une constante.

C'est pour ces raisons que l'on s'intéresse aux propriétés de l'estimateur $\widehat{\theta}_n$ dans une configuration très particulière, où la taille de l'échantillon n est suffisamment grande pour que l'on puisse utiliser des résultats de convergence (▶ chapitre 8), mais est supposée finie. Dans ce contexte précis, nous allons caractériser la distribution asymptotique de l'estimateur représentée par le symbole $\overset{asy}{\approx}$ (avec *asy* pour asymptotique).

$$\widehat{\theta}_n \overset{asy}{\approx} \text{loi asymptotique} \tag{9.69}$$

Le tableau 9.1 synthétise les différentes notions de distribution en fonction de la taille d'échantillon n, pour un estimateur $\widehat{\theta}_n$ convergent au sens fort[8], dont on ne connaît pas la loi exacte.

8 Dans le cas d'un estimateur convergent au sens faible, la dernière colonne devient : $\widehat{\theta}_n \overset{p}{\to} \theta$, convergence faible et $\widehat{\theta}_n$ est « presque » une constante.

▼ **Tableau 9.1 Résumé des propriétés d'un estimateur en fonction de n**

	Dimension n finie		Dimension n infinie
Taille n	n petit	n grand	$n \to \infty$
Résultat	$\widehat{\theta}_n \sim ?$	$\widehat{\theta}_n \overset{asy}{\approx}$ loi asymptotique	$\widehat{\theta}_n \overset{a.s.}{\to} \theta_0$
Interprétation	loi exacte inconnue	loi asymptotique	convergence forte
	$\widehat{\theta}_n$ est une v.a.r.	$\widehat{\theta}_n$ est une v.a.r. normale	$\widehat{\theta}_n$ est une constante

Note : le terme v.a.r. signifie variable aléatoire réelle. n désigne la taille de l'échantillon.

Comment déterminer la distribution asymptotique d'un estimateur ? La distribution asymptotique est généralement basée sur un résultat de **convergence en distribution**. Cela peut paraître paradoxal puisque la distribution asymptotique est valable pour une dimension n fixe alors que la convergence en distribution implique que n tende vers l'infini (▸ chapitre 8). Toutefois ce résultat de convergence en distribution ne porte pas directement sur l'estimateur $\widehat{\theta}_n$, mais sur une **variable transformée** qui dépend de la dimension n (qui définit la **vitesse de convergence**). L'idée est de déterminer une variable transformée de $\widehat{\theta}_n$ qui converge en loi vers une distribution non dégénérée, c'est-à-dire une distribution dont la variance ne tende ni vers 0, ni vers l'infini.

Exemple

Soit un n-échantillon de variables $(X_1,\ldots,X_n)$ i.i.d. telles que $\mathbb{E}(X_i) = \mu$ et $\mathbb{V}(X_i) = \sigma^2$, où l'espérance μ est un paramètre inconnu. On considère un estimateur $\widehat{\mu}_n$ du paramètre μ défini par la moyenne empirique :

$$\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{9.70}$$

Dans ce cas, la loi exacte de $\widehat{\mu}_n$ n'est pas connue puisque la loi des variables de l'échantillon $X_1,\ldots,X_n$ est elle-même inconnue. D'après la loi faible des grands nombres (théorème de Khintchine), nous savons que l'estimateur est convergent au sens faible :

$$\widehat{\mu}_n \overset{a.s.}{\to} \mu \tag{9.71}$$

La distribution de $\widehat{\mu}_n$ est donc dégénérée lorsque n tend vers l'infini. Mais, d'après le théorème central limite de Lindeberg-Levy, nous savons que la variable transformée $\sqrt{n}(\widehat{\mu}_n - \mu)$ converge en distribution vers une loi non dégénérée, puisque :

$$\underbrace{\sqrt{n}(\widehat{\mu}_n - \mu)}_{\text{variable transformée}} \overset{d}{\to} \underbrace{\mathcal{N}\left(0,\sigma^2\right)}_{\text{loi non dégénérée}} \tag{9.72}$$

À partir de ce résultat de convergence sur une variable transformée de $\widehat{\theta}_n$ dépendant de n, on déduit la distribution asymptotique de $\widehat{\theta}_n$ pour une valeur n fixe. Supposons que l'on obtienne un résultat de convergence en distribution pour $n \to \infty$, du type :

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \overset{d}{\to} \mathcal{N}(0,\Sigma) \tag{9.73}$$

On admet alors que pour une taille d'échantillon n très grande, mais finie ($n = 10\,000$ par exemple), on peut utiliser l'approximation suivante :

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \approx \mathcal{N}(0,\Sigma) \tag{9.74}$$

où le symbole $\approx$ signifie « *approximativement distribué selon* ». Puisque la dimension n est finie, on peut alors en déduire la distribution asymptotique de $\widehat{\theta}_n$ en ré-

arrangeant les termes de l'expression (9.74). Rappelons que si $aX \sim \mathcal{N}(b,c)$ alors $X \sim \mathcal{N}\left(b/a\ ,\ c/a^2\right)$. On en déduit que :

$$\widehat{\theta}_n - \theta_0 \approx \mathcal{N}\left(0, \frac{\Sigma}{n}\right) \tag{9.75}$$

De même si $X - a \sim \mathcal{N}(b,c)$ alors $X \sim \mathcal{N}(b - a\ ,\ c)$. Par conséquent la distribution asymptotique de l'estimateur $\widehat{\theta}_n$ est définie par :

$$\widehat{\theta}_n \overset{asy}{\approx} \mathcal{N}\left(\theta_0, \frac{\Sigma}{n}\right) \tag{9.76}$$

Exemple

Soit un n-échantillon de variables $(X_1,\ldots,X_n)$ i.i.d. telles que $\mathbb{E}(X_i) = \mu$ et $\mathbb{V}(X_i) = \sigma^2$, où l'espérance μ est un paramètre inconnu. On considère un estimateur $\widehat{\mu}_n$ du paramètre μ défini par la moyenne empirique, $\widehat{\mu}_n = n^{-1} \sum_{i=1}^{n} X_i$. D'après le théorème central limite de Lindeberg-Levy, nous savons que :

$$\sqrt{n}\,(\widehat{\mu}_n - \mu) \overset{d}{\to} \mathcal{N}\left(0,\sigma^2\right) \tag{9.77}$$

On en déduit la distribution asymptotique de l'estimateur $\widehat{\mu}_n$ pour une dimension n suffisament grande et finie :

$$\widehat{\mu}_n \overset{asy}{\approx} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \tag{9.78}$$

Dans de nombreux cas, les estimateurs que nous étudierons, convergent en distribution vers une loi normale (comme dans le cas de l'exemple précédent).

Définition 9.16

Un estimateur $\widehat{\theta}_n$ est **asymptotiquement normalement distribué** dès lors que :

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \overset{d}{\to} \mathcal{N}(0,\Sigma) \tag{9.79}$$

Sa **distribution asymptotique** est définie par :

$$\widehat{\theta}_n \overset{asy}{\approx} \mathcal{N}\left(\theta_0, \frac{\Sigma}{n}\right) \tag{9.80}$$

Pour être précis, il convient de ne pas confondre le résultat de convergence en distribution (9.79) qui porte sur une transformée de $\widehat{\theta}_n$, et la distribution asymptotique de $\widehat{\theta}_n$ (9.80), qui est la conséquence du résultat de convergence. Toutefois, parfois on utilise le terme de *distribution asymptotique* pour qualifier le résultat de convergence de l'équation (9.79). En relation avec le résultat précédent, nous pouvons à présent définir les concepts d'espérance et de variance asymptotiques :

Définition 9.17

Soit un estimateur $\widehat{\theta}_n$ convergent et asymptotiquement normalement distribué vérifiant :

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \overset{d}{\to} \mathcal{N}(0,\Sigma) \quad \text{ou} \quad \widehat{\theta}_n \overset{asy}{\approx} \mathcal{N}\left(\theta_0, \frac{\Sigma}{n}\right) \tag{9.81}$$

L'**espérance** et la **variance asymptotiques** de $\widehat{\theta}_n$ sont respectivement définies par :

$$\mathbb{E}_{asy}\left(\widehat{\theta}_n\right) = \theta_0 \qquad \mathbb{V}_{asy}\left(\widehat{\theta}_n\right) = \frac{\Sigma}{n} \tag{9.82}$$

On vérifie que, puisque l'estimateur est convergent, sa variance asymptotique tend vers 0 lorsque n tend vers l'infini :

$$\lim_{n\to\infty} \mathbb{V}_{asy}\left(\widehat{\theta}_n\right) = \lim_{n\to\infty} \frac{\Sigma}{n} = 0 \tag{9.83}$$

Reste à savoir comment obtenir un résultat de convergence en distribution pour un estimateur $\widehat{\theta}_n$ qui n'est pas simplement défini par la moyenne empirique. En général, la démarche est la suivante :

- **Étape 1.** On exprime l'estimateur $\widehat{\theta}_n$ comme une fonction $h(.)$ de la moyenne empirique $\overline{X}_n$ des variables de l'échantillon $X_1,\ldots,X_n$.

$$\widehat{\theta}_n = h\left(\overline{X}_n\right) \tag{9.84}$$

- **Étape 2.** On applique une version du théorème central limite afin de déterminer la convergence en distribution de la moyenne empirique $\overline{X}_n$. Dans le cas d'un échantillon i.i.d. avec $\mathbb{E}(X_i) = \mu$ et $\mathbb{V}(X_i) = \sigma^2$, le théorème central limite de Lindeberg-Levy nous permet d'obtenir :

$$\sqrt{n}\left(\overline{X}_n - \mu\right) \overset{d}{\to} \mathcal{N}(0,\Omega) \tag{9.85}$$

- **Étape 3.** On applique le théorème de Slutsky et/ou la méthode delta (► chapitre 8) pour déduire du résultat précédent la convergence en loi de l'estimateur $\widehat{\theta}_n = h\left(\overline{X}_n\right)$:

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \overset{d}{\to} \mathcal{N}(0,\Sigma) \tag{9.86}$$

  La forme de Σ est alors une fonction, plus ou moins compliquée, de Ω et éventuellement d'autres paramètres.

5 Estimation

Une fois que l'on dispose d'un « bon » estimateur $\widehat{\theta}$ (sans biais, efficace et convergent), la dernière étape consiste à l'appliquer à partir des données de l'échantillon $(x_1,\ldots,x_n)$ afin d'obtenir une **estimation** du paramètre ou du vecteur de paramètres θ. On distingue deux principales méthodes d'estimation[9] :

– l'estimation ponctuelle ;

– l'estimation par intervalle de confiance.

9 Il existe en fait une troisième méthode : l'estimation par densité. Plutôt que de donner une valeur (estimation ponctuelle) ou un intervalle de confiance pour estimer la valeur du paramètre θ, on fournit tout simplement la fonction de densité de l'estimateur $\widehat{\theta}$. Cette méthode est notamment utilisée pour la prévision (*density forecast*), mais la logique est la même pour l'estimation d'un paramètre. L'utilisateur de l'estimation (ou de la prévision) peut se faire une idée précise de l'incertitude autour de l'estimation (prévision) ponctuelle. Cette méthode est par exemple utilisée par la Banque d'Angleterre pour ses prévisions d'inflation (*fan charts*). Dans la pratique, la densité de l'estimateur est souvent estimée par des méthodes semi ou non-paramétriques.

5.1 Estimation ponctuelle

Dans la section 2.2, nous avons défini la notion d'**estimation ponctuelle**. Rappelons qu'il s'agit tout simplement de la réalisation $\widehat{\theta}(x)$ de l'estimateur $\widehat{\theta} = g(X_1,\ldots,X_n)$, obtenue à partir de la réalisation $(x_1,\ldots,x_n)$ de l'échantillon. Concrètement, obtenir une estimation revient tout simplement à appliquer la « formule » $\widehat{\theta} = g(X_1,\ldots,X_n)$ sur les données $(x_1,\ldots,x_n)$. Cette opération se fait habituellement à l'aide d'un logiciel d'économétrie ou de statistique (Eviews, Rats, Stata, Matlab, Scilab, etc.) ou d'un tableur (Excel, par exemple ; ▶ En pratique : Estimation et logiciel d'économétrie).

EN PRATIQUE

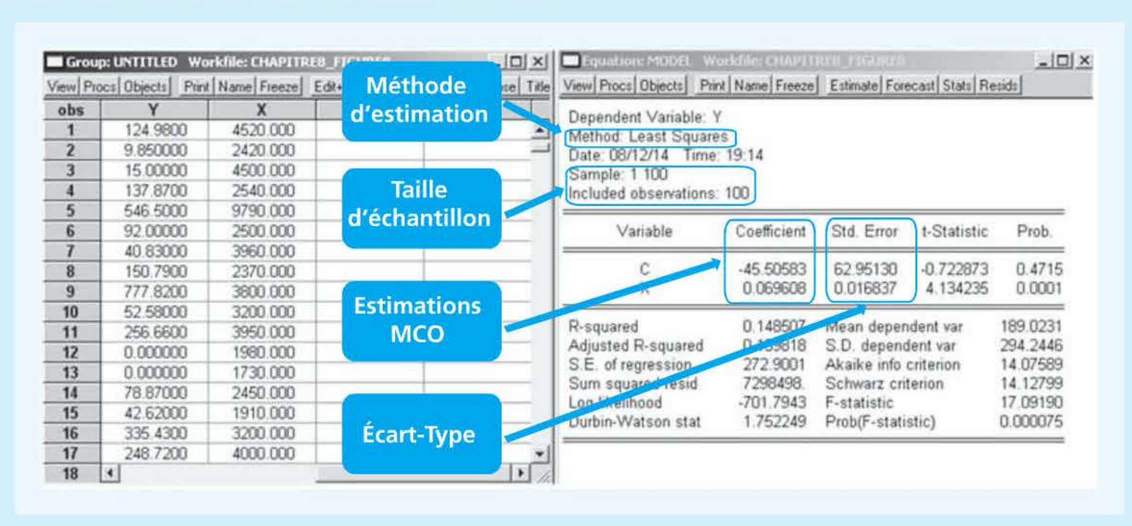
Estimation et logiciel d'économétrie

Sur la figure 9.8 est représenté un exemple de sortie du logiciel d'économétrie Eviews. Il s'agit d'un **modèle de régression linéaire** (▶ chapitre 2), dans lequel on explique le montant de dépenses mensuelles effectuées à l'aide d'une carte de crédit en dollars (variable Y_i) par le revenu du possesseur de la carte (variable X_i) selon l'équation :

$$Y_i = \alpha + \beta X_i + \varepsilon_i \qquad (9.87)$$

où les paramètres α et β sont inconnus et ε_i est un terme aléatoire d'erreur, de distribution inconnue, mais vérifiant $\mathbb{E}(\varepsilon_i) = 0$, $\mathbb{V}(\varepsilon_i) = \sigma^2$ et $\mathbb{E}(\varepsilon_i | x_i) = 0$. On dispose d'un échantillon $(x_i, y_i)_{i=1}^{n}$ de $n = 100$ individus pour lesquels on observe le revenu et le montant dépensé. Dans ce cas, la méthode d'estimation utilisée est celle des moindres carrés ordinaires (MCO ou *least squares* en anglais), présentée dans le chapitre 2. Soient $\widehat{\alpha}$ et $\widehat{\beta}$ les estimateurs des MCO des paramètres α et β. Sur la partie gauche de la figure 9.8, sont reportées les premières observations de l'échantillon. Sur la partie droite de la figure 9.8, on retrouve

- la méthode d'estimation utilisée (MCO) ;
- la taille de l'échantillon (100) ;
- les estimations $\widehat{\alpha}(x,y)$ et $\widehat{\beta}(x,y)$, *i.e.* les réalisations des estimateurs des MCO $\widehat{\alpha}$ et $\widehat{\beta}$ (−45,50 et 0,069) ;
- les écarts-types des estimateurs $\widehat{\alpha}$ et $\widehat{\beta}$ estimés à partir de leur distribution asymptotique (62,95 et 0,016).



▲ **Figure 9.8** Exemple de sortie de logiciel d'économétrie

5.2 Estimation par intervalle de confiance

Le principal inconvénient de l'estimation ponctuelle, c'est qu'elle ne rend pas compte de l'incertitude autour de l'estimation. Une façon de rendre compte de cette incertitude est de proposer un **intervalle de confiance** sur la valeur du paramètre θ.

Définition 9.18

Un **intervalle de confiance** sur le paramètre θ pour un niveau de confiance de $1-\alpha$ (ou un niveau de risque α), avec $\alpha \in]0,1[$, est un encadrement du type :

$$\Pr(A \leq \theta \leq B) = 1 - \alpha \tag{9.88}$$

où A et B sont des **variables aléatoires**, fonctions des variables de l'échantillon $X_1,\ldots,X_n$.

Une **réalisation** de cet intervalle de confiance est notée :

$$IC_{1-\alpha} = [a;b] \tag{9.89}$$

où a et b sont les réalisations respectives des variables A et B, obtenues à partir de la réalisation de l'échantillon $(x_1,\ldots,x_n)$.

Ainsi, on cherche à encadrer la valeur de θ (inconnue) par deux variables aléatoires A et B, telles que la probabilité que la valeur de θ soit comprise entre ces deux variables soit précisément égale à $1-\alpha$. Ces variables sont des fonctions des variables de l'échantillon. Dès lors, à partir de la réalisation $(x_1,\ldots,x_n)$ de l'échantillon, on peut déduire des réalisations des variables A et B, et une réalisation de l'intervalle, c'est-à-dire deux valeurs encadrant la vraie valeur θ pour un niveau de confiance de $1-\alpha$. Par exemple, si pour un échantillon $(x_1,\ldots x_n)$ et $\alpha = 5\ \%$, on obtient $IC_{0,95} = [1{,}2; 1{,}5]$, cela signifie que pour cette réalisation de l'échantillon (c'est-à-dire pour ces données), il y a 95 % de chances que la valeur de θ soit comprise entre 1,2 et 1,5.

Remarque : Il ne faut pas confondre l'**intervalle de confiance**, fondé sur des variables aléatoires et une probabilité (9.88), et sa **réalisation**, qui n'est qu'un segment défini par deux valeurs réelles (9.89). Ainsi, il convient d'**éviter** les notations du type :

$$\Pr(a \leq \theta \leq b) = 1 - \alpha \tag{9.90}$$

Par exemple, la notation $\Pr(1{,}2 \leq \theta \leq 1{,}5) = 0{,}95$ n'a pas de sens, car θ n'est pas une variable aléatoire. Il n'y aucune raison d'utiliser la probabilité dans ce cas puisque le paramètre θ est supposé constant : c'est ce qui distingue l'approche fréquentiste, utilisée ici, de l'approche Bayésienne de la statistique.

Comment obtenir un intervalle de confiance ? Il n'existe pas de méthode générale, mais la procédure suivante peut être utilisée dans de nombreux cas :

- **Étape 1.** On considère un estimateur $\widehat{\theta}$, sans biais et convergent, du paramètre θ. On cherche à caractériser soit (1) sa loi exacte, si cela est possible (ou celle d'une variable transformée), (2) soit sa loi asymptotique. Cette loi dépend nécessairement de θ puisque l'estimateur est sans biais, *i.e.* $\mathbb{E}(\widehat{\theta}) = \theta$.
- **Étape 2.** On transforme la variable $\widehat{\theta}$ de sorte à ce que la loi de la variable transformée ne dépende plus de θ, ni d'autres paramètres inconnus. Cette variable transformée dépend naturellement de θ (paramètre à estimer) et de $\widehat{\theta}$, mais elle ne doit

pas dépendre d'autres paramètres inconnus. Soit $h(\widehat{\theta},\theta)$ la variable transformée. On cherche à obtenir un résultat du type :

$$h(\widehat{\theta},\theta) \sim \text{loi connue (ne dépendant pas de paramètres inconnus)}$$

En général, on utilise ici une **z-transformée** du type :

$$z = \frac{\widehat{\theta} - \mathbb{E}(\widehat{\theta})}{\sqrt{\mathbb{V}(\widehat{\theta})}} = \frac{\widehat{\theta} - \theta}{\sqrt{\mathbb{V}(\widehat{\theta})}} \tag{9.91}$$

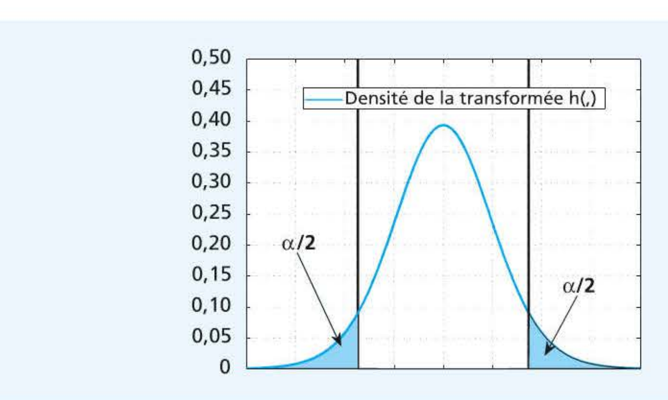
Dans certains cas, cette variable transformée dépend d'autres paramètres inconnus que θ. Il faut alors chercher à les remplacer par leurs estimateurs.

- **Étape 3.** À partir de la loi de la variable aléatoire transformée $h(\widehat{\theta},\theta)$, on construit un encadrement du type :

$$\Pr\left(\underbrace{c}_{\text{constante}} \leq \underbrace{h(\widehat{\theta},\theta)}_{\text{variable aléatoire}} \leq \underbrace{d}_{\text{constante}}\right) = 1 - \alpha \tag{9.92}$$

Pour cela, on cherche deux constantes réelles c et d, telles que que $d > c$ et que la distance $d - c$ soit la plus petite possible. Pour un intervalle symétrique, on peut obtenir les valeurs c et d de la façon suivante, comme l'illustre la figure 9.9 :

$$\Pr\left(h(\widehat{\theta},\theta) < c\right) = \frac{\alpha}{2} \qquad \Pr\left(h(\widehat{\theta},\theta) > d\right) = \frac{\alpha}{2} \tag{9.93}$$



▲ Figure 9.9 Intervalle de confiance

- **Étape 4.** En réaménageant les termes de cet encadrement, on cherche à construire un encadrement sur la valeur de θ, tel que :

$$\Pr\left(\underbrace{f\left(\widehat{\theta},c,d\right)}_{\text{variable aléatoire}} \leq \underbrace{\theta}_{\text{constante}} \leq \underbrace{g\left(\widehat{\theta},c,d\right)}_{\text{variable aléatoire}}\right) = \Pr\left(A \leq \theta \leq B\right) = 1 - \alpha \tag{9.94}$$

où $f(.)$ et $g(.)$ sont des fonctions, $A = f(\widehat{\theta},c,d)$ et $B = g(\widehat{\theta},c,d)$ sont des variables aléatoires qui dépendent de l'estimateur $\widehat{\theta}$ et donc implicitement des variables de l'échantillon $X_1,\ldots,X_n$. Elles dépendent en outre des constantes c ou d, suivant les transformations effectuées.

- **Étape 5.** À partir de la réalisation de l'échantillon $(x_1, \ldots, x_n)$ et de l'estimation ponctuelle $\widehat{\theta}(x)$, on déduit la réalisation de l'intervalle de confiance :

$$IC_{1-\alpha} = \left[f\left(\widehat{\theta}(x)\right); g\left(\widehat{\theta}(x)\right) \right] = [a; b] \tag{9.95}$$

Appliquons cette méthodologie dans le cadre de deux exemples. Le premier porte sur la construction d'un intervalle de confiance sur l'espérance d'une loi normale lorsque la variance est connue. Le second exemple porte sur la construction d'un intervalle de confiance sur l'espérance d'une loi normale lorsque la variance est inconnue.

Exemple

On souhaite estimer par intervalle de confiance l'espérance μ d'une variable aléatoire X suivant une loi normale de variance connue $\sigma^2 = 6{,}25$ à l'aide d'un échantillon de $n = 10$ variables $(X_1, \ldots, X_n)$ i.i.d. de même loi que X. On sait que la réalisation de la moyenne empirique $\overline{X}_n$ pour cet échantillon est égale à 4,3. On considère un niveau de risque $\alpha = 5\ \%$. Détaillons les étapes de la démarche.

- **Étape 1.** Nous savons que la moyenne empirique $\overline{X}_n$ est un estimateur sans biais et convergent (théorème de Khintchine) de l'espérance μ :

$$\overline{X}_n \overset{p}{\to} \mu \tag{9.96}$$

De plus, dans un échantillon N.i.d., la moyenne empirique a une distribution exacte normale :

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \tag{9.97}$$

- **Étape 2.** Construisons une variable transformée de $\overline{X}_n$ dont la loi ne dépend pas de paramètres inconnus. Ici, il suffit d'utiliser la z-transformée :

$$\frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{\mathbb{V}(\overline{X}_n)}} = \frac{\overline{X}_n - \mu}{\sigma/n} \sim \underbrace{\mathcal{N}(0,1)}_{\text{loi ne dépendant pas de } \theta} \tag{9.98}$$

Remarque : On aurait pu utiliser le résultat $\overline{X}_n - \mu \sim \mathcal{N}\left(0, \sigma^2\right)$, puisque la variance σ^2 est connue.

- **Étape 3.** On construit un encadrement du type :

$$\Pr\left(c \leq \frac{\overline{X}_n - \mu}{\sigma/n} \leq d \right) = 1 - \alpha \tag{9.99}$$

Les constantes c et d sont telles que :

$$\Pr\left(\frac{\overline{X}_n - \mu}{\sigma/n} < c \right) = \Phi(c) = \frac{\alpha}{2} \Longleftrightarrow c = \Phi^{-1}\left(\frac{\alpha}{2}\right) \tag{9.100}$$

$$\Pr\left(\frac{\overline{X}_n - \mu}{\sigma/n} > d \right) = 1 - \Pr\left(\frac{\overline{X}_n - \mu}{\sigma/n} \leq d \right) = \frac{\alpha}{2} \Longleftrightarrow d = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \tag{9.101}$$

où $\Phi(.)$ désigne la fonction de répartition de la loi normale centrée réduite.

- **Étape 4.** De l'encadrement précédent, on déduit que :

$$\Pr\left(\Phi^{-1}\left(\frac{\alpha}{2}\right) \leq \frac{\overline{X}_n - \mu}{\sigma/n} \leq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \right) = 1 - \alpha \tag{9.102}$$

$$\Longleftrightarrow \Pr\left(\frac{\sigma}{n}\Phi^{-1}\left(\frac{\alpha}{2}\right) - \overline{X}_n \leq -\mu \leq \frac{\sigma}{n}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) - \overline{X}_n \right) = 1 - \alpha \tag{9.103}$$

En multipliant par −1 les termes de ces inégalités, il faut penser à inverser les bornes de l'encadrement. Ainsi, on obtient un intervalle de confiance à 95 % du type :

$$\Pr\left(\underbrace{\overline{X}_n - \frac{\sigma}{n}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)}_{\text{variable } A} \leq \mu \leq \underbrace{\overline{X}_n - \frac{\sigma}{n}\Phi^{-1}\left(\frac{\alpha}{2}\right)}_{\text{variable } B}\right) = 1-\alpha \qquad (9.104)$$

La loi normale étant symétrique, $\Phi^{-1}(\alpha/2) = -\Phi^{-1}(1-\alpha/2)$, on peut réécrire cet intervalle de confiance sous la forme :

$$\Pr\left(\overline{X}_n - \frac{\sigma}{n}\Phi^{-1}\left(1-\frac{\alpha}{2}\right) \leq \mu \leq \overline{X}_n + \frac{\sigma}{n}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right) = 1-\alpha \qquad (9.105)$$

- **Étape 5.** Pour un niveau de risque $\alpha = 0{,}05$, une taille d'échantillon $n = 10$, une variance $\sigma^2 = 6{,}25$ et une réalisation de la moyenne empirique $\overline{x}_n = 4{,}3$, on obtient une réalisation de l'intervalle de confiance égale à :

$$\begin{aligned}
IC_{0.95} &= \left[\overline{x}_n - \frac{\sigma}{n}\Phi^{-1}\left(1-\frac{\alpha}{2}\right); \overline{x}_n - \frac{\sigma}{n}\Phi^{-1}\left(\frac{\alpha}{2}\right)\right] \\
&= \left[4{,}3 - \frac{\sqrt{6{,}25}}{10} \times \Phi^{-1}(0{,}975); 4{,}3 - \frac{\sqrt{6{,}25}}{10} \times \Phi^{-1}(0{,}025)\right] \\
&= \left[4{,}3 - \frac{2{,}5}{10} \times (1{,}96); 4{,}3 - \frac{2{,}5}{10} \times (-1{,}96)\right] \\
&= [3{,}81; 4{,}79] \qquad (9.106)
\end{aligned}$$

Exemple

On considère le même exemple que précédemment (échantillon normal, $n = 10$, $\overline{x}_n = 4{,}3$), en supposant cette fois-ci que la variance des variables X_i n'est pas connue. On suppose en outre que la réalisation de la variance empirique corrigée $S_n^2 = (n-1)^{-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$ est égale à 6,76 sur cet échantillon. Construisons un intervalle de confiance à à 95 % sur $\mathbb{E}(X_i) = \mu$.

- **Étape 1.** Nous savons que la moyenne empirique $\overline{X}_n$ est un estimateur sans biais et convergent (théorème de Khintchine) de l'espérance μ. Sa distribution exacte est :

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \qquad (9.107)$$

- **Étape 2.** Construisons une variable transformée de $\overline{X}_n$ dont la loi ne dépend pas de paramètres inconnus. Dans ce cas, on ne peut plus utiliser la z-transformée, car cette dernière dépend d'un paramètre inconnu, σ :

$$\frac{\overline{X}_n - \mu}{\underbrace{\sigma}_{\text{paramètre inconnu}}/n} \sim \mathcal{N}(0,1) \qquad (9.108)$$

Nous allons donc remplacer σ^2 par un estimateur convergent, à savoir la variance empirique corrigée S_n^2. Nous savons que dans un échantillon normal :

$$\left(\frac{n-1}{\sigma^2}\right)S_n^2 \sim \chi^2(n-1) \qquad (9.109)$$

Par ailleurs, on peut démontrer que les deux variables définies par les équations (9.108) et (9.109) sont indépendantes. Rappelons que si X et Y sont deux variables indépendantes telles que $X \sim \mathcal{N}(0,1)$ et $Y \sim \chi^2(v)$, alors la variable $Z = X/\sqrt{(Y/v)}$ suit une distribution

de Student à v degrés de liberté, notée $t(v)$. Dans notre cas, nous pouvons définir une variable telle que :

$$\frac{\frac{\overline{X}_n-\mu}{\sigma/n}}{\sqrt{\left(\frac{n-1}{\sigma^2}\right)\frac{S_n^2}{(n-1)}}} = \left(\frac{\overline{X}_n-\mu}{\sigma/n}\right)\frac{\sigma}{S n} = \frac{\overline{X}_n-\mu}{S_n/n} \sim t(n-1) \qquad (9.110)$$

On observe que cette variable ne dépend pas de paramètres inconnus (hormis μ, le paramètre que l'on souhaite estimer) et sa loi ne dépend pas, elle aussi, de paramètres inconnus.

- **Étape 3.** On construit un encadrement du type :

$$\Pr\left(c \leq \frac{\overline{X}_n-\mu}{S_n/n} \leq d\right) = 1-\alpha \qquad (9.111)$$

Les constantes c et d sont telles que :

$$\Pr\left(\frac{\overline{X}_n-\mu}{S_n/n} < c\right) = F_{n-1}(c) = \frac{\alpha}{2} \Longleftrightarrow c = F_{n-1}^{-1}\left(\frac{\alpha}{2}\right) \qquad (9.112)$$

$$\Pr\left(\frac{\overline{X}_n-\mu}{S_n/n} > d\right) = 1-\Pr\left(\frac{\overline{X}_n-\mu}{S_n/n} \leq d\right) = \frac{\alpha}{2} \Longleftrightarrow d = F_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right) \qquad (9.113)$$

où $F_{n-1}(.)$ désigne la fonction de répartition de la loi de Student à $n-1$ degrés de liberté.

- **Étape 4.** De l'encadrement précédent, on déduit un encadrement sur μ :

$$\Pr\left(\underbrace{\overline{X}_n-\frac{S_n}{n}F_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right)}_{\text{variable } A} \leq \mu \leq \underbrace{\overline{X}_n-\frac{S_n}{n}F_{n-1}^{-1}\left(\frac{\alpha}{2}\right)}_{\text{variable } B}\right) = 1-\alpha \qquad (9.114)$$

La loi de Student étant symétrique, $F_{n-1}^{-1}(\alpha/2) = -F_{n-1}^{-1}(1-\alpha/2)$, on peut réécrire cet intervalle de confiance sous la forme :

$$\Pr\left(\overline{X}_n-\frac{S_n}{n}F_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right) \leq \mu \leq \overline{X}_n+\frac{S_n}{n}F_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right)\right) = 1-\alpha \qquad (9.115)$$

- **Étape 5.** Pour un niveau de risque $\alpha = 0{,}05$, une taille d'échantillon $n = 10$, une réalisation de la variance empirique corrigée $s_n^2 = 6{,}76$ et une réalisation de la moyenne empirique $\overline{x}_n = 4{,}3$, on obtient une réalisation de l'intervalle de confiance égale à :

$$\begin{aligned} IC_{0.95} &= \left[\overline{x}_n-\frac{s_n}{n}F_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right); \overline{x}_n-\frac{s_n}{n}F_{n-1}^{-1}\left(\frac{\alpha}{2}\right)\right] \\ &= \left[4{,}3-\frac{\sqrt{6{,}76}}{10}\times F_9^{-1}(0{,}975)\,; 4{,}3-\frac{\sqrt{6{,}76}}{10}\times F_9^{-1}(0{,}025)\right] \\ &= \left[4{,}3-\frac{2{,}6}{10}\times(2{,}2622)\,; 4{,}3-\frac{2{,}6}{10}\times(-2{,}2622)\right] \\ &= [3{,}6240; 4{,}9760] \end{aligned} \qquad (9.116)$$

où $F_9(.)$ désigne la fonction de répartition de la loi de Student à 9 degrés de liberté.



3 questions à

Ekaterina Sborets

Senior Risk Analyst, Lloyds Banking Group (Londres)



Quel est votre parcours professionnel et votre mission actuelle à la Lloyds ?

À l'issue de mon master d'Econométrie et de Statistique appliquée à l'Université d'Orléans, j'ai été embauchée par BNP Paribas Personal Finance au sein du Centre de Scoring. En 2014, j'ai rejoint la Lloyds Banking Group à Londres en tant que Senior Risk Analyst. Actuellement, mon rôle consiste à développer des modèles statistiques qui permettent d'analyser et de prévoir le comportement des clients de la Banque afin de définir des stratégies concernant la politique d'octroi des prêts ou des cartes de crédit, l'ouverture des comptes courants, les étapes de recouvrement des créances impayées, etc.

Quelle est l'importance de la phase de constitution de l'échantillon dans votre activité ?

La phase de constitution d'un échantillon est essentielle dans tout travail de modélisation. L'échantillonnage intervient à deux niveaux : lors de la phase d'estimation des paramètres du modèle de risque et lors de la phase de validation de ce modèle. On constitue généralement deux échantillons (une base d'apprentissage et une base de test) en vérifiant l'affectation d'une proportion spécifique de « bons » et de « mauvais » individus dans la base totale.
Nous utilisons des méthodes d'échantillonnage aléatoires pour éviter le phénomène de sur-apprentissage sur des niches de population. Il faut être conscient que la question de la volumétrie des données est de plus en plus importante. C'est pourquoi, l'échantillonnage résulte aussi d'un arbitrage entre des impératifs statistiques (représentativité de l'échantillon, taille suffisante) et des contraintes opérationnelles (réduction des coûts et du temps de calcul).

Quels sont les méthodes d'estimation que vous utilisez pour estimer les paramètres de ces modèles de risque ?

Suivant le modèle retenu, les paramètres sont estimés par des méthodes paramétriques (maximum de vraisemblance, moindres carrés ordinaires) ou semi-paramétriques (méthode des moments généralisés). On utilise aussi parfois des approches non paramétriques (estimateurs kernel, régression locales polynomiales). ■

Les points **clés**

- Un échantillon aléatoire est une collection de variables aléatoires.
- Un estimateur est une variable aléatoire.
- Une estimation ponctuelle correspond à la réalisation de l'estimateur obtenue pour un échantillon (réalisation).
- La distribution à distance finie ou distribution exacte d'un estimateur est valable pour toute taille de l'échantillon.
- La distribution asymptotique d'un estimateur est valable pour un échantillon de très grande taille.
- Un estimateur est non biaisé si son espérance correspond à la valeur du paramètre à estimer.
- Un estimateur est convergent s'il converge en probabilité vers la vraie valeur du paramètre.
- Un estimateur est asymptotiquement normalement distribué si sa distribution asymptotique est normale.
- Un estimateur est efficace si sa variance atteint la borne de Cramer-Rao.
- Un intervalle de confiance est un encadrement de la vraie valeur du paramètre par deux variables aléatoires.

ÉVALUATION

▶ Corrigés sur www.dunod.com

QCM

Pour chacune des questions suivantes, indiquez si les affirmations sont vraies ou fausses (il peut y avoir plusieurs réponses vraies pour chaque question).

1 Propriétés d'un estimateur

a. Un estimateur sans biais est efficace.

b. Un estimateur sans biais est convergent.

c. Un estimateur convergent au sens fort est nécessairement convergent au sens faible.

d. Un estimateur convergent est sans biais.

e. Un estimateur efficace est sans biais.

2 Variance empirique

On considère un n -échantillon de variables N.i.d.

a. La variance empirique est un estimateur de la variance.

b. La variance empirique corrigée est un estimateur sans biais de la variance.

c. La variance empirique corrigée a une distribution exacte du khi-deux.

d. Une transformée de la variance empirique corrigée admet une distribution du khi-deux à n degrés de libertés.

e. La variance empirique corrigée est définie par $S_n^2 = n^{-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$.

3 Comparaison d'estimateurs

Soient deux estimateurs sans biais $\widehat{\theta}_1$ et $\widehat{\theta}_2$.

a. L'estimateur $\widehat{\theta}_1$ est préféré à $\widehat{\theta}_2$ si sa variance est plus faible.

b. L'estimateur $\widehat{\theta}_1$ est efficace au sens FDCR si sa variance est plus faible que celle de $\widehat{\theta}_2$.

c. Si l'estimateur $\widehat{\theta}_1$ est efficace au sens FDCR, alors sa variance est plus faible ou égale à celle de $\widehat{\theta}_2$.

d. Si l'estimateur $\widehat{\theta}_1$ est optimal alors il est efficace au sens de la borne FDCR.

e. Un estimateur $\widehat{\theta}_1$ convergent est efficace.

4 Intervalle de confiance

a. Un intervalle de confiance est un encadrement sur la vraie valeur du paramètre par deux variables aléatoires.

b. Un intervalle de confiance est fondé sur un estimateur sans biais et convergent.

c. Plus la variance de l'estimateur est faible, plus l'amplitude de la réalisation de l'intervalle de confiance sera faible.

d. Le niveau de risque d'un intervalle de confiance est généralement plus faible que le niveau de confiance.

e. Un intervalle de confiance est un segment de deux valeurs a et b.

Sujets d'examen

5 Estimation et loi de Rayleigh (HEC Lausanne, 2013)

Soient deux variables aléatoires réelles X_1 et X_2 indépendantes et distribuées chacune selon une loi $\mathcal{N}\left(0,\sigma^2\right)$. On admet que la variable aléatoire transformée Y définie par la relation :

$$Y = \sqrt{\left(X_1^2 + X_2^2\right)} \quad (9.117)$$

suit une *loi de Rayleigh* de paramètre σ^2, avec $\sigma > 0$. On admet que cette variable Y a pour fonction de densité :

$$f_Y\left(y;\sigma^2\right) = \frac{y}{\sigma^2}\exp\left(-\frac{y^2}{2\sigma^2}\right) \quad \forall y \in [0, +\infty[\quad (9.118)$$

On suppose que le paramètre σ^2 est inconnu et on cherche à l'estimer à partir d'un n-échantillon $(Y_1,\ldots,Y_n)$ de variables i.i.d. de même loi que Y. On considère un estimateur $\widehat{\sigma^2}$ du paramètre σ^2 défini par :

$$\widehat{\sigma^2} = \frac{1}{2n}\sum_{i=1}^{n} Y_i^2 \quad (9.119)$$

1. Quelle est la loi de la variable Y^2/σ^2 ? En déduire la valeur des moments $\mathbb{E}\left(Y^2\right)$ et $\mathbb{V}\left(Y^2\right)$.

2. Montrez que l'estimateur $\widehat{\sigma}^2$ est sans biais.

3. Montrez que l'estimateur $\widehat{\sigma}^2$ est convergent.

4. Montrez que l'estimateur $\widehat{\sigma}^2$ vérifie :

$$\sqrt{n}\left(\widehat{\sigma}^2 - \sigma^2\right) \xrightarrow{d} \mathcal{N}\left(0,\sigma^4\right) \qquad (9.120)$$

5. Quelle est la variance asymptotique de l'estimateur $\widehat{\sigma}^2$?

6. Montrez que l'estimateur $\widehat{\sigma}^2$ est efficace au sens de la borne FDCR. On admet que $\mathrm{I}_n(\theta) = n/\sigma^4$, où σ est la vraie valeur du paramètre de la loi de Rayleigh.

6 Loi exacte et loi asymptotique (Université d'Orléans, 2012)

On considère un n-échantillon de variables $(Z_1,\ldots,Z_n)$ i.i.d. de même loi que Z, où Z suit une loi normale centrée réduite. On considère une variable D_n telle que :

$$D_n = \sum_{i=1}^{n} Z_i^2 \qquad (9.121)$$

1. Quelle est la loi exacte de la variable D_n ?

2. On admet que les variables $\left(Z_1^2,\ldots,Z_n^2\right)$ sont i.i.d. Par application du théorème central limite, déterminez la loi asymptotique de la variable transformée :

$$\sqrt{n}\left(\frac{D_n}{n} - 1\right) \qquad (9.122)$$

3. On suppose que l'on dispose d'un échantillon de taille $n = 100$. Calculez la probabilité $\Pr(D_n > 118{,}49)$ en utilisant (*i*) la loi exacte et (*ii*) la loi asymptotique.

7 Comparaison d'estimateurs (Université d'Orléans, 2011)

On considère une variable aléatoire continue X distribuée selon une loi de probabilité telle que $\mathbb{E}(X) = \theta$ et $\mathbb{V}(X) = \theta - \theta^2$ où θ est un paramètre inconnu vérifiant $\theta \in\ ]0,1[$. Soit un n-échantillon $(X_1,\ldots,X_n)$ i.i.d. de même loi que X. Soient $\widehat{\theta}_1$ et $\widehat{\theta}_2$ deux estimateurs du paramètre θ respectivement définis par :

$$\widehat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \widehat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \qquad (9.123)$$

1. Montrer que les estimateurs $\widehat{\theta}_1$ et $\widehat{\theta}_2$ sont sans biais.

2. Montrer que les estimateurs $\widehat{\theta}_1$ et $\widehat{\theta}_2$ sont convergents (au sens de la convergence en probabilité). On admettra que $\mathbb{V}\left(X^2\right) = 2\theta^2 - 2\theta^4$.

3. Peut-on déterminer quel est l'estimateur le plus précis ?

4. Quelles sont les lois asymptotiques des estimateurs $\widehat{\theta}_1$ et $\widehat{\theta}_2$?

8 Estimation (d'après HEC Lausanne, 2014)

On considère un échantillon $\{X_1,\ldots,X_n\}$ de variables aléatoires continues i.i.d. de même loi que X, où X est définie sur le support $X(\Omega) = [0,c]$ et admet une fonction de densité égale à :

$$f_X(x;\theta) = \frac{1}{\theta c^{1/\theta}} x^{\frac{1-\theta}{\theta}} \quad \forall x \in X(\Omega) \qquad (9.124)$$

On suppose que la borne c est connue et que le paramètre θ est un paramètre positif inconnu que l'on cherche à estimer. On admet que l'estimateur du maximum de vraisemblance (▸ chapitre 10) du paramètre θ est défini par :

$$\widehat{\theta} = \ln(c) - \frac{1}{n}\sum_{i=1}^{n} \ln(X_i) \qquad (9.125)$$

On admet que :

$$\mathbb{E}(\ln(X_i)) = \ln(c) - \theta \qquad (9.126)$$

1. Montrer que l'estimateur $\widehat{\theta}$ est sans biais.

2. Montrer que l'estimateur $\widehat{\theta}$ est convergent.

Chapitre 10

Lorsque vous sollicitez un crédit à la consommation sur Internet auprès d'une banque, vous remplissez généralement un formulaire en ligne et la banque vous donne immédiatement une réponse de principe (sous réserve de produire par la suite un certain nombre de documents). Cette réponse automatique est issue d'un modèle statistique que l'on appelle un modèle de score d'octroi.

Par comparaison de vos caractéristiques socio-individuelles (salaire, âge, type d'emploi, etc.) et de celles de clients passés, ce modèle de scoring permet à la banque d'évaluer votre niveau de risque et de vous donner une réponse immédiate quant à l'octroi ou non du prêt. Ces modèles de scoring sont généralement des modèles paramétriques et leurs paramètres sont presque toujours estimés par la méthode du maximum de vraisemblance.



Maximum de vraisemblance

Plan

Pré-requis

→ **Connaître** les différentes notions de convergence (▸ chapitre 8).
→ **Connaître** la notion d'estimateur (▸ chapitre 9).

Objectifs

→ **Comprendre** la notion de vraisemblance.
→ **Savoir utiliser** l'estimateur du maximum de vraisemblance.
→ **Savoir analyser** les propriétés de l'estimateur du maximum de vraisemblance.
→ **Comprendre** les notions de score et de matrice hessienne.
→ **Comprendre** les différentes notions de matrices d'information de Fisher.

La procédure du maximum de vraisemblance est une **méthode d'estimation** (▶ chapitre 9). Il s'agit donc d'une méthode statistique qui permet de dériver la forme fonctionnelle (ou la « formule ») d'un estimateur particulier : l'**estimateur du maximum de vraisemblance**. Le principe de cette méthode est extrêmement simple : on part de l'hypothèse que la variable d'intérêt suit une certaine **distribution paramétrique**, *i.e.* une distribution caractérisée par un nombre fini de paramètres. Ces paramètres sont inconnus et l'on cherche à les estimer. On utilise pour cela un échantillon (collection de variables aléatoires) pour lequel on dispose d'une réalisation, c'est-à-dire d'un ensemble d'observations. Si les variables de l'échantillon sont discrètes, on construit la probabilité jointe d'apparition des données de l'échantillon. Dans le cas continu, on construit la densité jointe associée à ces observations. Cette probabilité jointe ou cette densité jointe correspond à la **vraisemblance de l'échantillon**. La vraisemblance est une fonction des observations et des paramètres inconnus de la distribution : elle mesure la plausibilité des données observées conditionnellement à une hypothèse de distribution sur la variable d'intérêt et à une valeur des paramètres. Le principe du maximum de vraisemblance consiste alors à déterminer la valeur des paramètres qui rend l'échantillon observé le plus vraisemblable. Dit autrement, la forme de l'estimateur du maximum de vraisemblance est déterminée par la maximisation de la vraisemblance de l'échantillon.

Cette méthode d'estimation est sans doute la plus utilisée en statistique et en économétrie. Les paramètres de la plupart des modèles non-linéaires considérés de nos jours en marketing, en finance, en gestion des risques (scoring bancaire), en assurance, etc., sont estimés par maximum de vraisemblance. Une des raisons de ce succès est que, sous des hypothèses relativement générales dites **hypothèses de régularité**, l'estimateur du maximum de vraisemblance présente de très bonnes propriétés. On peut notamment montrer que cet estimateur est sans biais, efficace et convergent (▶ chapitre 9). Il est par ailleurs asymptotiquement normalement distribué. Cette dernière propriété est particulièrement remarquable. Quelle que soit la distribution supposée de la variable d'intérêt (Poisson, exponentielle, Student, khi-deux, etc.), l'estimateur du maximum de vraisemblance des paramètres associés à cette distribution particulière converge toujours vers une distribution asymptotique normale. C'est pourquoi cette méthode d'estimation est aujourd'hui disponible dans tous les logiciels d'économétrie et dans certains tableurs.

1 Principe du maximum de vraisemblance

Dans cette section, nous allons introduire le principe de l'**estimation par maximum de vraisemblance** à partir d'un exemple.

On considère un n-échantillon $(X_1,\ldots,X_n)$ de variables aléatoires discrètes, positives et i.i.d. On suppose que ces variables admettent une distribution de Poisson (▶ chapitre 7) de paramètre $\theta > 0$, où θ est un paramètre inconnu que l'on souhaite estimer. La fonction de masse des variables X_i, pour $i = 1,\ldots,n$, est la suivante :

$$\Pr(X_i = x) = \frac{\exp(-\theta)\,\theta^x}{x!} \quad \forall x \in \mathbb{N} \tag{10.1}$$

On dispose d'une **réalisation de l'échantillon** (observations), notée $(x_1,\ldots,x_n)$. Par exemple, pour $n = 10$, on observe (5,0,1,1,0,3,2,3,4,1). Quelle est la probabilité d'observer précisément cette réalisation de l'échantillon ? Cette probabilité est égale à :

$$\Pr((X_1 = x_1) \cap \ldots \cap (X_n = x_n)) \tag{10.2}$$

Puisque les variables X_i sont indépendantes, cette **probabilité jointe** est égale au produit des **probabilités marginales** :

$$\Pr((X_1 = x_1) \cap \ldots \cap (X_n = x_n)) = \prod_{i=1}^{n} \Pr(X_i = x_i) \tag{10.3}$$

Si l'on suppose que les variables X_i admettent une distribution de Poisson de paramètre θ, on obtient :

$$\Pr((X_1 = x_1) \cap \ldots \cap (X_n = x_n)) = \prod_{i=1}^{n} \frac{\exp(-\theta)\,\theta^{x_i}}{x_i!} = \exp(-n\theta)\,\frac{\theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \tag{10.4}$$

Cette probabilité jointe est une fonction de θ (le paramètre inconnu) et de l'échantillon $(x_1,\ldots,x_n)$: elle correspond à la fonction[1] de **vraisemblance de l'échantillon**. On note cette vraisemblance sous la forme suivante :

$$L_n(\theta; x_1,\ldots,x_n) = \Pr((X_1 = x_1) \cap \ldots \cap (X_n = x_n)) \tag{10.5}$$

avec dans notre cas :

$$L_n(\theta; x_1,\ldots,x_n) = \exp(-n\theta) \times \theta^{\sum_{i=1}^{n} x_i} \times \frac{1}{\prod_{i=1}^{n} x_i!} \tag{10.6}$$

Exemple

Supposons que la taille d'échantillon soit égale à 10 et que l'on ait une réalisation de l'échantillon (données) égale à (5,0,1,1,0,3,2,3,4,1), alors :

$$L_n(\theta; x_1,\ldots,x_n) = \Pr((X_1 = x_1) \cap \ldots \cap (X_n = x_n)) = \frac{\exp(-10\theta) \times \theta^{20}}{207\,360} \tag{10.7}$$

L'**intuition** de l'estimation par maximum de vraisemblance consiste à déterminer la valeur du paramètre θ qui maximise cette probabilité d'apparition de l'échantillon $(x_1,\ldots,x_n)$, c'est-à-dire qui **maximise la vraisemblance de l'échantillon**. La figure 10.1 représente la fonction $L_n(\theta; x_1,\ldots,x_n)$ pour l'échantillon (5,0,1,1,0,3,2,3,4,1) et pour différentes valeurs positives du paramètre θ. On constate immédiatement que cette fonction atteint son maximum pour une valeur de θ égale à 2.

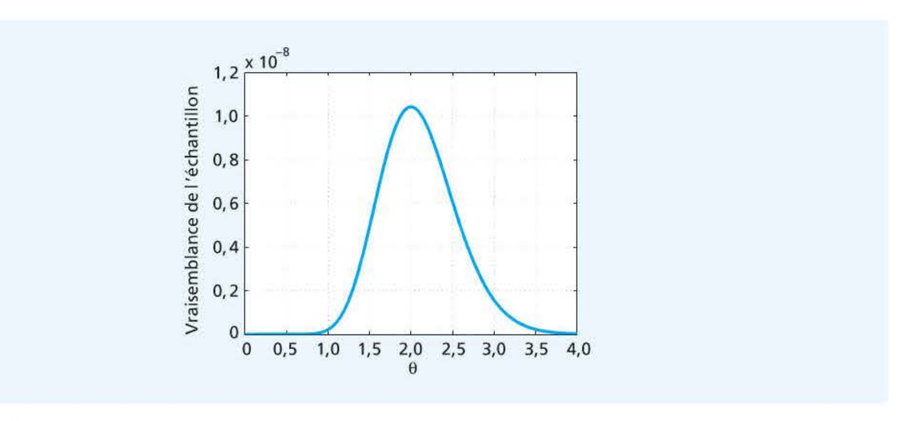
On peut vérifier analytiquement que cette valeur correspond bien au maximum de la fonction de vraisemblance. Pour cela, considérons le programme de maximisation[2] suivant :

$$\widehat{\theta} = \underset{\theta \in \mathbb{R}^+}{\arg\max}\, L_n(\theta; x_1,\ldots,x_n) \tag{10.8}$$

Dans la mesure où il est souvent plus facile de considérer des sommes que des produits et que maximiser la fonction de vraisemblance est équivalent à maximiser le

1 Cette fonction dépend de deux arguments, *i.e.* le paramètre et l'échantillon, que l'on sépare dans les notations par un point-virgule, étant donnée leur nature très différente.

2 Le terme arg max signifie l'argument qui maximise. En effet, $\widehat{\theta}$ est défini comme l'argument de la fonction de vraisemblance qui maximise cette fonction.



▲ **Figure 10.1** **Vraisemblance de l'échantillon**

logarithme de cette fonction, on préfère en général maximiser la **log-vraisemblance**. Le programme devient alors le suivant :

$$\widehat{\theta} = \underset{\theta \in \mathbb{R}^+}{\arg\max} \ln L_n(\theta; x_1, \ldots, x_n) \tag{10.9}$$

$$\ln L_n(\theta; x_1, \ldots, x_n) = -n\theta + \ln(\theta) \sum_{i=1}^{n} x_i - \ln\left(\prod_{i=1}^{n} x_i!\right) \tag{10.10}$$

La condition nécessaire (CN) de ce programme est la suivante :

$$\text{CN} : \left.\frac{\partial \ln L_n(\theta; x_1, \ldots, x_n)}{\partial \theta}\right|_{\widehat{\theta}} = -n + \frac{1}{\widehat{\theta}} \sum_{i=1}^{n} x_i = 0 \tag{10.11}$$

On en déduit immédiatement que la valeur qui maximise la log-vraisemblance correspond à la moyenne empirique :

$$\widehat{\theta} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{10.12}$$

À présent, il convient de vérifier que l'on a bien un maximum. Pour ce faire, on considère la condition suffisante (CS) du programme de maximisation :

$$\text{CS} : \left.\frac{\partial^2 \ln L_n(\theta; x_1, \ldots, x_n)}{\partial \theta^2}\right|_{\widehat{\theta}} = -\frac{1}{\widehat{\theta}^2} \sum_{i=1}^{n} x_i < 0 \tag{10.13}$$

Cette quantité étant négative, on a bien un maximum. Notons que le maximum de la fonction de log-vraisemblance (équation (10.12)) est une fonction des réalisations de l'échantillon $x_1, \ldots, x_n$. C'est donc une quantité déterministe (réalisation) qui correspond à la réalisation de la moyenne empirique.

Comme nous l'avons vu dans le chapitre 9, il convient de ne pas confondre un estimateur (variable aléatoire) et sa réalisation (quantité déterministe). Étant donnés les résultats précédents, on peut définir l'**estimateur du maximum de vraisemblance** (MV) de la façon suivante :

$$\text{Estimateur du MV} : \widehat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{10.14}$$

L'estimateur $\widehat{\theta}$ du paramètre θ est une variable aléatoire définie comme une fonction des variables aléatoires de l'échantillon $X_1, \ldots, X_n$. Pour distinguer l'estimateur $\widehat{\theta}$ de sa

réalisation, on note cette dernière $\widehat{\theta}(x)$. Dans notre cas :

$$\text{Réalisation : } \widehat{\theta}(x) = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (10.15)$$

Pour l'échantillon (5,0,1,1,0,3,2,3,4,1), on vérifie que l'on obtient $\widehat{\theta}(x) = 2$.

Remarque : Dans cet exemple introductif, nous avons considéré des variables X_i discrètes. Bien évidemment, la méthode du maximum de vraisemblance s'applique aussi à des **variables aléatoires continues**. La seule différence est que, dans ce cas, la vraisemblance ne s'analyse plus comme une probabilité jointe. En effet, pour une variable continue $\Pr(X_i = x_i) = 0$, puisque la probabilité d'être en un point particulier est nulle. La vraisemblance de l'échantillon correspond alors à la **fonction de densité de la loi jointe** des variables $X_1,\ldots,X_n$ évaluée au point $(x_1,\ldots,x_n)$:

$$L_n(\theta; x_1,\ldots,x_n) = f_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta) \qquad (10.16)$$

FOCUS

Programme de maximisation

De façon générale, la fonction de vraisemblance dépend de deux arguments : le paramètre et la réalisation de l'échantillon. L'estimateur du maximum de vraisemblance est obtenu en maximisant cette fonction par rapport au paramètre. Considérons une fonction f à deux variables θ et x, notée $f(\theta; x)$ où $\theta \in \mathbb{R}$ représente le paramètre et x l'échantillon. La maximisation de cette fonction par rapport au premier argument correspond au **programme de maximisation** suivant :

$$\theta^* = \arg\max_{\theta\in\mathbb{R}} f(\theta; x) \qquad (10.17)$$

La **condition nécessaire** de ce programme consiste à annuler la dérivée partielle de la fonction $f(\theta; x)$ par rapport à son premier argument θ, notée à l'aide du symbole ∂. Il est important de préciser que la dérivée partielle $\partial f(\theta; x)/\partial\theta$ est elle-même une fonction des variables θ et x. Par exemple, si $f(\theta; x) = \theta^2 x$, alors $\partial f(\theta; x)/\partial\theta = 2\theta x$. Notons $g(\theta; x)$ cette fonction. On cherche donc la valeur θ^* telle que cette dérivée partielle soit nulle :

$$g(\theta^*; x) = \left.\frac{\partial f(\theta; x)}{\partial\theta}\right|_{\theta^*, x} = \left.\frac{\partial f(\theta; x)}{\partial\theta}\right|_{\theta^*} = 0 \qquad (10.18)$$

La notation avec une barre verticale signifie que l'on considère la fonction à gauche de cette barre (en l'occurrence $\partial f(\theta; x)/\partial\theta$) et que l'on évalue cette fonction au point situé à droite de la barre verticale, c'est-à-dire le couple $(\theta^*; x)$. Par souci de simplification, le point d'évaluation est souvent noté θ^* puisque le second argument de la fonction est toujours le même et correspond à l'échantillon.

La **condition suffisante** permet de vérifier que la solution θ^* est bien un maximum. Pour cela, il suffit de vérifier que la dérivée seconde de $f(\theta; x)$ par rapport à θ, évaluée au point θ^*, est négative. On note cette dérivée seconde sous la forme $h(\theta; x) = \partial^2 f(\theta; x)/\partial\theta^2$. Notons que le carré apparaît sur le symbole de la dérivée partielle ∂ et sur l'argument par rapport auquel on dérive : θ^2 signifie donc que l'on dérive deux fois par rapport à θ. Par exemple, si $f(\theta; x) = \theta^2 x$, alors $\partial^2 f(\theta; x)/\partial\theta^2 = 2x$. De façon générale, on cherche à vérifier que :

$$h(\theta^*; x) = \left.\frac{\partial^2 f(\theta; x)}{\partial\theta^2}\right|_{\theta^*} = \left.\frac{\partial g(\theta; x)}{\partial\theta}\right|_{\theta^*} < 0 \qquad (10.19)$$

2 Fonction de vraisemblance

Nous commencerons par présenter le concept de fonction de vraisemblance dans le cadre d'un problème avec un seul paramètre à estimer, puis nous étendrons cette définition au cas avec plusieurs paramètres et à la notion de modèle économétrique.

2.1 Définitions

Soit X une variable aléatoire (discrète ou continue) définie sur un univers probabilisé $(X(\Omega), \mathcal{F}, \Pr)$, dont la loi de probabilité est caractérisée par une fonction de densité ou une fonction de masse notée $f_X(x;\theta)$, $\forall x \in X(\Omega)$. Cette fonction dépend d'un **paramètre inconnu**, noté θ, avec $\theta \in \Theta \subset \mathbb{R}$ où Θ désigne l'ensemble des valeurs possibles pour ce paramètre. Afin d'estimer θ, on dispose d'un n-échantillon $(X_1,\ldots,X_n)$ de variables i.i.d. de même loi que X. La réalisation de cet échantillon est notée $(x_1,\ldots,x_n)$ ou x en abrégé.

Remarque : La méthode d'estimation du maximum de vraisemblance suppose que l'on connaisse la loi de la variable X ou de façon équivalente, la loi des variables de l'échantillon. Plus précisément, on connaît la forme de la fonction de densité (ou de masse) de X, mais cette forme dépend d'un paramètre inconnu. Il y a donc une sorte de « *pari* » sur la distribution de X. Mieux vaut ne pas se tromper...

Ainsi, afin d'appliquer la méthode du maximum de vraisemblance, il est absolument nécessaire de « postuler » une distribution paramétrique pour la variable d'intérêt, c'est-à-dire une fonction de densité ou de masse paramétrée par un ou plusieurs paramètres inconnus.

Exemple

On suppose que la durée de vie d'un équipement peut être représentée par une variable aléatoire continue et positive D admettant une distribution exponentielle d'intensité $1/\theta$ où θ est un paramètre réel positif. Sa fonction de densité est définie par :

$$f_D(d;\theta) = \frac{1}{\theta}\exp\left(-\frac{d}{\theta}\right) \quad \forall d \in \mathbb{R}^+ \tag{10.20}$$

Sous ces hypothèses, nous pouvons déterminer la **vraisemblance de l'échantillon**. Dans la section 1, nous avons vu que la vraisemblance est définie par la densité ou la probabilité jointe associée aux réalisations de l'échantillon. Si les variables $X_1,\ldots,X_n$ sont indépendantes, cette densité ou cette probabilité jointe peut s'écrire comme le produit des densités ou des probabilités marginales.

Définition 10.1

La fonction de **vraisemblance de l'échantillon** $(x_1,\ldots,x_n)$ est définie par :

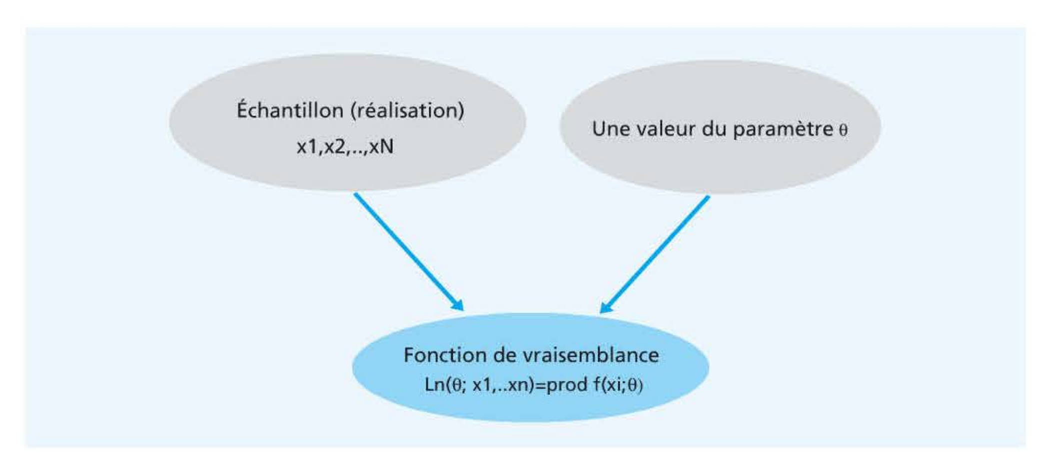
$$L_n : \Theta \times X(\Omega)^n \to \mathbb{R}^+ \tag{10.21}$$

$$(\theta; x_1,\ldots,x_n) \longmapsto L_n(\theta; x_1,\ldots,x_n) = \prod_{i=1}^{n} f_X(x_i;\theta) \tag{10.22}$$

La vraisemblance étant définie comme un produit de fonctions de densité ou de probabilités (fonction de masse), cette quantité est nécessairement *positive*.

Remarque : Il est important de spécifier ce à quoi se rapporte la fonction de vraisemblance : il s'agit soit de la vraisemblance d'un *échantillon*, soit de la vraisemblance d'une *observation*, etc. Il convient d'éviter, dans la mesure du possible, les expressions du type « fonction de vraisemblance » ou « vraisemblance ».

Comme l'illustre la figure 10.2, la fonction de vraisemblance d'un échantillon dépend de deux arguments : le paramètre θ et la réalisation de l'échantillon $(x_1,\ldots,x_n)$. Ces deux arguments sont des constantes déterministes : la vraisemblance de l'échantillon est donc une quantité déterministe (valeur constante).



▲ **Figure 10.2** **Fonction de vraisemblance d'un échantillon**

Définition 10.2

La fonction de **log-vraisemblance de l'échantillon** $(x_1,\ldots,x_n)$ est définie par :

$$\ell_n : \Theta \times X(\Omega)^n \to \mathbb{R} \tag{10.23}$$

$$(\theta; x_1,\ldots,x_n) \longmapsto \ell_n(\theta; x_1,\ldots,x_n) = \sum_{i=1}^{n} \ln f_X(x_i;\theta) \tag{10.24}$$

La fonction de log-vraisemblance, contrairement à la vraisemblance, peut être positive ou négative. Bien évidemment, on vérifie que :

$$\ell_n(\theta; x_1,\ldots,x_n) = \ln L_n(\theta; x_1,\ldots,x_n) \tag{10.25}$$

Exemple

On considère un n-échantillon $(D_1,\ldots,D_n)$ de variables aléatoires continues, positives et i.i.d. On suppose que les variables D_i admettent une distribution exponentielle $\mathcal{E}xp(1/\theta)$, où $\theta > 0$ est un paramètre inconnu. La fonction de densité des variables D_i est définie par :

$$f_D(d_i;\theta) = \frac{1}{\theta}\exp\left(-\frac{d_i}{\theta}\right) \quad \forall d_i \in \mathbb{R}^+ \tag{10.26}$$

Puisque les variables D_i sont indépendantes, la vraisemblance associée à l'échantillon $(d_1,\ldots,d_n)$ est définie par :

$$L_n(\theta; d_1,\ldots,d_n) = \prod_{i=1}^{n} f_D(d_i;\theta) = \prod_{i=1}^{n} \frac{1}{\theta}\exp\left(-\frac{d_i}{\theta}\right) \quad (10.27)$$

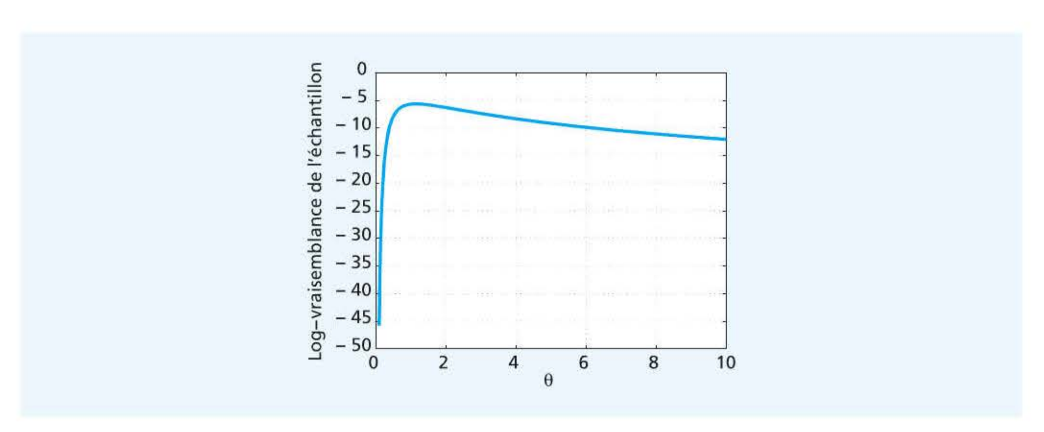
$$= \theta^{-n}\exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} d_i\right) \quad (10.28)$$

La log-vraisemblance de l'échantillon est définie par :

$$\ell_n(\theta; d_1,\ldots,d_n) = \sum_{i=1}^{n} \ln f_D(d_i;\theta) = \sum_{i=1}^{n}\left(-\ln(\theta) - \frac{d_i}{\theta}\right) \quad (10.29)$$

$$= -n\ln(\theta) - \frac{1}{\theta}\sum_{i=1}^{n} d_i \quad (10.30)$$

La figure 10.3 représente la fonction de log-vraisemblance obtenue pour un échantillon {3,6952; 0,0597; 0,0876; 1,4457; 0,4456} de taille $n = 5$.



▲ **Figure 10.3 Log-vraisemblance de l'échantillon**

Remarque : Afin de simplifier les notations, on note parfois les fonctions de vraisemblance et de log-vraisemblance d'un échantillon de la façon suivante :

$$L_n(\theta; x) \equiv L(\theta; x_1,\ldots,x_n) \equiv L_n(\theta) \quad (10.31)$$

$$\ell_n(\theta; x) \equiv \ln L_n(\theta; x) \equiv \ln L(\theta; x_1,\ldots,x_n) \equiv \ln L_n(\theta) \quad (10.32)$$

Il est aussi possible de définir la vraisemblance (ou la log-vraisemblance) d'une observation particulière.

Définition 10.3

La vraisemblance et la log-vraisemblance associées à **une observation** x_i, pour $i \in \{1,\ldots,n\}$, sont respectivement définies par :

$$L_i(\theta; x) = f_X(x_i;\theta) \qquad \ell_i(\theta; x) = \ln f_X(x_i;\theta) \quad (10.33)$$

Par construction, ces quantités vérifient :

$$L_n(\theta;x) = \prod_{i=1}^{n} L_i(\theta;x), \qquad \ell_n(\theta;x) = \sum_{i=1}^{n} \ell_i(\theta;x) \tag{10.34}$$

Reprenons l'exemple précédent d'un échantillon de variables distribuées selon une loi exponentielle.

Exemple

On considère un n-échantillon $(D_1,\ldots,D_n)$ de variables aléatoires continues et positives. On suppose que ces variables sont i.i.d. et suivent une distribution exponentielle $\mathcal{E}xp(1/\theta)$ avec $\theta > 0$. Soit $(d_1,\ldots,d_n)$ la réalisation de cet échantillon. La vraisemblance et la log-vraisemblance associées à l'observation d_i, $\forall i = 1,\ldots,n$, sont respectivement définies par :

$$L_i(\theta;d_i) = f_D(d_i;\theta) = \frac{1}{\theta}\exp\left(-\frac{d_i}{\theta}\right) \tag{10.35}$$

$$\ell_i(\theta;d_i) = \ln(f_D(d_i;\theta)) = -\ln(\theta) - \frac{d_i}{\theta} \tag{10.36}$$

Si $d_1 = 2$, on a $L_1(\theta;d_1) = (1/\theta)\exp(-2/\theta)$ et $\ell_1(\theta;d_1) = -\ln(\theta) - 2/\theta$.

2.2 Extension au cas avec plusieurs paramètres

Bien souvent, la distribution de la variable d'intérêt X ne dépend pas uniquement d'un seul paramètre, mais d'un ensemble de k paramètres. On définit alors un **vecteur de paramètres**, noté θ, de dimension $k \times 1$ tel que :

$$\theta = \begin{pmatrix} \theta_1 \\ \ldots \\ \theta_k \end{pmatrix} \tag{10.37}$$

Exemple

Soit une variable Y telle que $Y \sim \mathcal{N}(\mu,\sigma^2)$ alors :

$$f_Y(y;\theta) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \qquad \forall y \in \mathbb{R} \tag{10.38}$$

où μ et σ^2 sont des paramètres inconnus. On pose $k = 2$ et un vecteur θ défini par :

$$\theta = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix} \tag{10.39}$$

Le fait de considérer un vecteur de paramètres ne change rien aux définitions des fonctions de vraisemblance et de log-vraisemblance associées à l'échantillon $(x_1,\ldots,x_n)$ ou à l'observation x_i. Reprenons l'exemple précédent.

Exemple

Soit un n-échantillon $(Y_1,\ldots,Y_n)$ N.i.d.(μ,σ^2) et $(y_1,\ldots,y_n)$ sa réalisation. Si l'on définit un vecteur de paramètres $\theta = (\mu\ \sigma^2)^\top$, alors les fonctions de vraisemblance et de log-

vraisemblance associées à l'échantillon sont respectivement définies par :

$$L_n(\theta; y) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right) \tag{10.40}$$

$$= \left(\sigma^2 2\pi\right)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right) \tag{10.41}$$

$$\ell_n(\theta; y) = \ln L_n(\theta; y) = -\frac{n}{2}\ln\left(\sigma^2\right) - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2 \tag{10.42}$$

2.3 Modèle et vraisemblance conditionnelle

Il est possible d'utiliser la méthode du maximum de vraisemblance pour estimer les paramètres d'un **modèle économétrique**. Un modèle économétrique peut être défini comme une relation théorique entre une variable Y dite endogène (ou dépendante) et une ou plusieurs variables X dites exogènes (ou explicatives).

$$Y = g(X; \theta) + \varepsilon \tag{10.43}$$

où θ est un vecteur de paramètres, $g(.)$ une fonction de lien et ε est un terme d'erreur, supposé aléatoire. Dans ce cas, il convient de considérer la **distribution conditionnelle** de Y sachant que les variables X prennent une certaine valeur. C'est à partir de cette distribution conditionnelle que nous allons déterminer la vraisemblance de l'échantillon $(y_i,x_i)_{i=1}^n$: on parle alors de **vraisemblance conditionnelle**.

Considérons le problème général. Soient deux variables aléatoires continues[3], notées X et Y. On suppose que la variable Y admet une distribution conditionnelle sachant que $X = x$, caractérisée par une fonction de densité conditionnelle notée $f_{Y|x}(y;\theta)$, $\forall y \in Y(\Omega) \subseteq \mathbb{R}$. Le paramètre $\theta \in \Theta \subset \mathbb{R}$ est inconnu et l'on cherche à l'estimer. Pour ce faire, on dispose d'un n-échantillon $(X_i,Y_i)_{i=1}^n$ et une réalisation $(x_i,y_n)_{i=1}^n$. Les variables Y_i peuvent être dépendantes pour $i = 1,\ldots,n$, mais l'on suppose qu'elles sont indépendantes conditionnellement à $X_i = x_i$.

La fonction de densité (ou fonction de masse) associée à la distribution conditionnelle de Y sachant $X = x$ peut s'écrire sous différentes formes :

$$f_{Y|x}(y;\theta) \equiv f_{Y|X}(y|\, x;\theta) \equiv f_Y(y|\, X = x;\theta) \equiv f_Y(y|\, X = x) \tag{10.44}$$

Sous ces hypothèses, on peut définir la vraisemblance et la log-vraisemblance conditionnelles associées à l'échantillon.

Définition 10.4

La fonction de **vraisemblance conditionnelle** et la **log-vraisemblance conditionnelle de l'échantillon** $(y_i,x_i)_{i=1}^n$ sont respectivement définies par :

$$L_n(\theta; y|\, x) = \prod_{i=1}^{n} f_{Y|X}(y_i|\, x_i;\theta), \qquad \ell_n(\theta; y|\, x) = \sum_{i=1}^{n} \ln f_{Y|X}(y_i|\, x_i;\theta) \tag{10.45}$$

3 On peut aussi envisager le cas où les variables X et Y sont des variables discrètes. Il suffit alors de considérer la fonction de masse conditionnelle (probabilité conditionnelle) en lieu et place de la densité conditionnelle afin de définir la vraisemblance.

où $f_{Y|X}(y_i|x_i;\theta)$ désigne la densité conditionnelle de la variable Y_i sachant $X_i = x_i$.

Exemple

On considère un modèle de régression linéaire tel que :

$$Y_i = X_i\beta + \varepsilon_i \quad (10.46)$$

où X_i est une variable explicative et β un paramètre. On suppose que le terme d'erreur ε_i est i.i.d. avec $\varepsilon_i \sim \mathcal{N}(0,\sigma^2)$. Sous ces hypothèses, la *distribution conditionnelle* de Y_i sachant $X_i = x_i$ est une distribution normale telle que :

$$Y_i|\, x_i \sim \mathcal{N}(x_i\beta,\sigma^2) \quad (10.47)$$

En effet, si $X_i = x_i$ alors $Y_i = x_i\beta + \varepsilon_i$ est la somme d'un terme constant ($x_i\beta$) et d'une variable normalement distribuée (ε_i). La densité conditionnelle de Y_i est donc :

$$f_{Y|X}(y_i|\,\boldsymbol{x}_i;\boldsymbol{\theta}) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(y_i - \boldsymbol{x}_i\boldsymbol{\beta})^2}{2\sigma^2}\right) \quad (10.48)$$

où $\theta = (\beta,\sigma^2)^\top$ est un vecteur de paramètres de dimension 2×1. Conditionnellement à $X_i = x_i$, les variables Y_i sont définies comme la somme de termes constants mais spécifiques à chaque observation ($x_i\beta$) et de termes i.i.d. (ε_i). Donc, les variables Y_i sont indépendantes conditionnellement à $X_i = x_i$, même si elles ne sont pas identiquement distribuées puisque $x_i\beta \neq x_j\beta$ pour $i \neq j$. Sous l'hypothèse d'indépendance, la vraisemblance conditionnelle de l'échantillon s'écrit :

$$L_n(\boldsymbol{\theta};\, y|\, x) = \prod_{i=1}^{n} f_{Y|X}(y_i|\,\boldsymbol{x}_i;\boldsymbol{\theta}) \quad (10.49)$$

$$= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(y_i - \boldsymbol{x}_i\boldsymbol{\beta})^2}{2\sigma^2}\right) \quad (10.50)$$

$$= \left(\sigma^2 2\pi\right)^{-n/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i\boldsymbol{\beta})^2\right) \quad (10.51)$$

La log-vraisemblance conditionnelle de l'échantillon est alors égale à :

$$\ell_n(\boldsymbol{\theta};\, y|\, x) = -\frac{n}{2}\ln\left(\sigma^2\right) - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i\boldsymbol{\beta})^2 \quad (10.52)$$

3 Estimateur du maximum de vraisemblance

Nous commencerons par présenter la notion d'estimateur de maximum de vraisemblance dans le cas où le paramètre θ est un scalaire, puis nous étendrons ces définitions au cas d'un vecteur de paramètres.

3.1 Définitions

Avant de définir l'estimateur du maximum de vraisemblance comme la quantité qui maximise la fonction de log-vraisemblance, il convient de s'assurer que le paramètre θ est **identifiable** à partir de cette fonction.

Définition 10.5

Le paramètre θ est **identifiable** (ou estimable) pour l'échantillon $x_1,\ldots,x_n$, si pour toutes valeurs θ^* et θ telles que $\theta^* \neq \theta$, les lois jointes des variables $(x_1, \ldots, x_n)$ sont différentes.

Tous les problèmes que nous considérerons dans cet ouvrage sont identifiables. Sous cette hypothèse, on peut définir l'estimateur du maximum de vraisemblance comme suit.

Définition 10.6

L'**estimateur du maximum de vraisemblance** $\widehat{\theta}$ du paramètre $\theta \in \Theta$ est la solution du problème de maximisation suivant :

$$\widehat{\theta} = \underset{\theta \in \Theta}{\arg\max}\ \ell_n(\theta; x) \tag{10.53}$$

De façon équivalente, on peut considérer le programme de maximisation de la vraisemblance $L_n(\theta; x)$. Mais il est souvent plus simple de maximiser la log-vraisemblance que la vraisemblance d'un échantillon.

Remarque : Rappelons qu'il convient de ne pas confondre l'estimateur $\widehat{\theta}$, qui est une variable aléatoire, et sa réalisation $\widehat{\theta}(x)$ qui est une constante. Puisque la log-vraisemblance dépend de la réalisation de l'échantillon $(x_1,\ldots,x_n)$, l'argument qui maximise cette fonction dépend lui aussi de cette réalisation. Ainsi au sens strict, le programme de maximisation devrait donc s'écrire sous la forme suivante :

$$\widehat{\theta}(x) = \underset{\theta \in \Theta}{\arg\max}\ \ell_n(\theta; x) \tag{10.54}$$

La résolution de ce programme permet d'obtenir l'**estimation** $\widehat{\theta}(x)$ associée aux données $x_1,\ldots,x_n$. De cette estimation, l'on déduit ensuite la forme fonctionnelle de l'estimateur $\widehat{\theta}$ exprimée comme une fonction des variables aléatoires $X_1,\ldots,X_n$. Toutefois, afin de simplifier les notations, nous utiliserons $\widehat{\theta}$ à la place de $\widehat{\theta}(x)$ dans le programme d'optimisation et dans les conditions nécessaires et suffisantes. Au-delà des notations, il convient de bien faire la différence entre les deux concepts.

La résolution du programme de maximisation de la log-vraisemblance, qui définit l'estimateur du maximum de vraisemblance, requiert de calculer la dérivée première et la dérivée seconde de cette fonction par rapport au paramètre θ. Ces dérivées correspondent respectivement au **gradient** et à la **hessienne**.

Définition 10.7

Le **gradient** de l'échantillon, noté $g_n(\theta; x)$, correspond à la dérivée partielle *première* de la fonction de log-vraisemblance de l'échantillon par rapport au paramètre θ :

$$g_n(\theta; x) = \frac{\partial \ell_n(\theta; x)}{\partial \theta} \tag{10.55}$$

Définition 10.8

La **hessienne** de l'échantillon, notée $H_n(\theta; x)$, correspond à la dérivée partielle *seconde* de la fonction de log-vraisemblance de l'échantillon par rapport au paramètre θ :

$$H_n(\theta; x) = \frac{\partial^2 \ell_n(\theta; x)}{\partial \theta^2} = \frac{\partial g_n(\theta; x)}{\partial \theta} \tag{10.56}$$

La condition nécessaire du programme de maximisation de la log-vraisemblance correspond à l'équation de log-vraisemblance.

Définition 10.9

On appelle **équation de log-vraisemblance** l'équation associée à la condition nécessaire du programme de maximisation de la log-vraisemblance :

$$\text{CN} : g_n\left(\widehat{\theta}; x\right) = \left.\frac{\partial \ell_n(\theta; x)}{\partial \theta}\right|_{\widehat{\theta}} = 0 \tag{10.57}$$

où $g_n(\theta; x)$ désigne le gradient associé à l'échantillon $x_1,\ldots,x_n$.

Ainsi, le gradient évalué au point $\widehat{\theta}$ (réalisation) doit être nul. La résolution de cette équation en $\widehat{\theta}$ permet d'obtenir l'estimation du maximum de vraisemblance en fonction des réalisations de l'échantillon (données) $x_1,\ldots,x_n$. De cette forme fonctionnelle, on déduira ensuite l'estimateur du maximum de vraisemblance. Mais avant cela, il convient de s'assurer que la solution $\widehat{\theta}$ est un maximum en vérifiant la condition suffisante du programme de maximisation.

Définition 10.10

La **condition suffisante** (CS) du programme de maximisation de la log-vraisemblance consiste à vérifier que la hessienne évaluée au point $\widehat{\theta}$ est négative :

$$\text{CS} : H_n\left(\widehat{\theta}; x\right) = \left.\frac{\partial^2 \ell_n(\theta; x)}{\partial \theta^2}\right|_{\widehat{\theta}} = \left.\frac{\partial g_n(\theta; x)}{\partial \theta}\right|_{\widehat{\theta}} < 0 \tag{10.58}$$

Appliquons ces définitions dans le cadre de deux problèmes d'estimation : le premier exemple concerne l'estimation d'un paramètre d'une loi discrète tandis que le second exemple porte sur une loi continue.

Exemple

On considère une variable aléatoire discrète X à valeurs dans $\mathbb{N}^*$, supposée suivre une *loi géométrique* de paramètre θ, avec $\theta \in]0,1[$. On rappelle que la fonction de densité de X est définie par (▶ chapitre 7) :

$$f_X(x; \theta) = \theta \times (1-\theta)^{x-1} \qquad \forall x \in \{1,2,3,\ldots\} \tag{10.59}$$

Soit un n-échantillon $(X_1,\ldots,X_n)$ où les variables X_i sont i.i.d. de même loi que X. Puisque les variables X_i sont indépendantes, la log-vraisemblance de l'échantillon $(x_1,\ldots,x_n)$ est égale à :

$$\ell_n(\theta;x) = \sum_{i=1}^{n} \ln f_X(x_i;\theta) = n\ln(\theta) + \sum_{i=1}^{n}(x_i-1)\ln(1-\theta) \qquad (10.60)$$

L'estimateur du maximum de vraisemblance $\widehat{\theta}$ est la solution du programme :

$$\widehat{\theta} = \underset{\theta\in]0,1[}{\arg\max}\ \ell_n(\theta;x) \qquad (10.61)$$

Le gradient et la hessienne de l'échantillon sont respectivement définis par :

$$g_n(\theta;x) = \frac{\partial \ell_n(\theta;x)}{\partial\theta} = \frac{n}{\theta} - \frac{1}{1-\theta}\sum_{i=1}^{n}(x_i-1) \qquad (10.62)$$

$$H_n(\theta;x) = \frac{\partial^2 \ell_n(\theta;x)}{\partial\theta^2} = -\frac{n}{\theta^2} - \frac{1}{(1-\theta)^2}\sum_{i=1}^{n}(x_i-1) \qquad (10.63)$$

La condition nécessaire du programme de maximisation s'écrit alors :

$$\text{CN}: g_n\left(\widehat{\theta};x\right) = \left.\frac{\partial \ell_n(\theta;x)}{\partial\theta}\right|_{\widehat{\theta}} = \frac{n}{\widehat{\theta}} - \frac{1}{1-\widehat{\theta}}\sum_{i=1}^{n}(x_i-1) = 0 \qquad (10.64)$$

En réarrangeant les termes, il vient :

$$\widehat{\theta} = n\left(\sum_{i=1}^{n} x_i\right)^{-1} = \frac{1}{\overline{x}_n} \qquad (10.65)$$

On vérifie que cette solution est un maximum :

$$H_n\left(\widehat{\theta};x\right) = \left.\frac{\partial^2 \ell_n(\theta;x)}{\partial\theta^2}\right|_{\widehat{\theta}} = -\frac{n}{\widehat{\theta}^2} - \frac{1}{(1-\widehat{\theta})^2}\sum_{i=1}^{n}(x_i-1) \qquad (10.66)$$

Puisque $\sum_{i=1}^{n} x_i = n/\widehat{\theta}$, cette expression peut se réécrire sous la forme :

$$H_n\left(\widehat{\theta};x\right) = -\frac{n}{\widehat{\theta}^2} - \frac{1}{(1-\widehat{\theta})^2}\left(\frac{n}{\widehat{\theta}} - n\right) \qquad (10.67)$$

$$= -\frac{n}{\widehat{\theta}^2} - \frac{n}{\widehat{\theta}(1-\widehat{\theta})} \qquad (10.68)$$

$$= -\frac{n}{\widehat{\theta}^2(1-\widehat{\theta})} < 0 \qquad (10.69)$$

Nous avons bien un maximum. Par conséquent, l'estimateur du maximum de vraisemblance du paramètre θ correspond à l'inverse de la moyenne empirique :

$$\widehat{\theta} = n\left(\sum_{i=1}^{n} X_i\right)^{-1} = \frac{1}{\overline{X}_n} \qquad (10.70)$$

Sa réalisation (estimation du maximum de vraisemblance) est égale à :

$$\widehat{\theta}(x) = n\left(\sum_{i=1}^{n} x_i\right)^{-1} = \frac{1}{\overline{x}_n} \qquad (10.71)$$

Exemple

Soit X une variable aléatoire réelle, continue, positive et caractérisée par une fonction de densité $f_X\left(x;\sigma^2\right)$ telle que :

$$f_X\left(x;\sigma^2\right) = \exp\left(-\frac{x^2}{2\sigma^2}\right)\frac{x}{\sigma^2} \quad \forall x \in \mathbb{R}^+ \tag{10.72}$$

où σ^2 est un paramètre inconnu. Afin d'estimer ce paramètre, on dispose d'un n-échantillon $(X_1,\ldots,X_n)$ de variables i.i.d. de même loi que X. Déterminons l'estimateur du maximum de vraisemblance du paramètre σ^2. Puisque les variables X_i sont indépendantes, la log-vraisemblance de l'échantillon $(x_1,\ldots,x_n)$ est définie par :

$$\ell_n\left(\sigma^2;x\right) = \sum_{i=1}^{n} \ln f_X\left(x_i;\sigma^2\right) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} \ln(x_i) - n\ln\left(\sigma^2\right) \tag{10.73}$$

L'estimateur du maximum de vraisemblance $\widehat{\sigma}^2$ est la solution du programme :

$$\widehat{\sigma}^2 = \underset{\sigma^2\in\mathbb{R}^+}{\arg\max}\ \ell_n\left(\sigma^2;x\right) \tag{10.74}$$

Le gradient et la hessienne de l'échantillon sont respectivement définis par :

$$g_n\left(\sigma^2;x\right) = \frac{\partial \ell_n\left(\sigma^2;x\right)}{\partial\sigma^2} = \frac{1}{2\sigma^4}\sum_{i=1}^{n} x_i^2 - \frac{n}{\sigma^2} \tag{10.75}$$

$$H_n\left(\sigma^2;x\right) = \frac{\partial^2 \ell_n\left(\sigma^2;x\right)}{\partial\sigma^4} = -\frac{1}{\sigma^6}\sum_{i=1}^{n} x_i^2 + \frac{n}{\sigma^4} \tag{10.76}$$

Conseil : Afin d'éviter les erreurs dans la dérivation par rapport à σ^2, une solution consiste à effectuer un changement de variable $\theta = \sigma^2$, puis à dériver deux fois la fonction de log-vraisemblance par rapport à θ. La condition nécessaire du programme de maximisation s'écrit alors :

$$\text{CN} : g_n\left(\widehat{\sigma}^2;x\right) = \left.\frac{\partial \ell_n\left(\sigma^2;x\right)}{\partial\sigma^2}\right|_{\widehat{\sigma}^2} = \frac{1}{2\widehat{\sigma}^4}\sum_{i=1}^{n} x_i^2 - \frac{n}{\widehat{\sigma}^2} = 0 \tag{10.77}$$

On en déduit que :

$$\widehat{\sigma}^2 = \frac{1}{2n}\sum_{i=1}^{n} x_i^2 \tag{10.78}$$

On vérifie que cette solution est un maximum :

$$H_n\left(\widehat{\sigma}^2;x\right) = \left.\frac{\partial^2 \ell_n\left(\sigma^2;x\right)}{\partial\sigma^4}\right|_{\widehat{\sigma}^2} = -\frac{1}{\widehat{\sigma}^6}\sum_{i=1}^{n} x_i^2 + \frac{n}{\widehat{\sigma}^4} \tag{10.79}$$

Puisque $2n\widehat{\sigma}^2 = \sum_{i=1}^{n} x_i^2$, cette expression peut se réécrire sous la forme suivante :

$$H_n\left(\widehat{\sigma}^2;x\right) = -\frac{2n\widehat{\sigma}^2}{\widehat{\sigma}^6} + \frac{n}{\widehat{\sigma}^4} = -\frac{n}{\widehat{\sigma}^4} < 0 \tag{10.80}$$

Nous avons bien un maximum. Par conséquent, l'estimateur du maximum de vraisemblance du paramètre σ^2 est défini par :

$$\widehat{\sigma}^2 = \frac{1}{2n}\sum_{i=1}^{n} X_i^2 \tag{10.81}$$

Sa réalisation (estimation du maximum de vraisemblance) est égale à :

$$\widehat{\sigma}^2(x) = \frac{1}{2n}\sum_{i=1}^{n} x_i^2 \tag{10.82}$$

3.2 Extension au cas avec plusieurs paramètres

Lorsque l'on considère un **vecteur de paramètres** $\theta = (\theta_1,\ldots,\theta_k)^\top$, la définition de l'estimateur du maximum de vraisemblance demeure inchangée. La seule différence est que, dans ce cas, le gradient est un vecteur de dimension $k \times 1$ et la hessienne est une matrice de dimension $k \times k$.

Les **équations de vraisemblance** correspondent alors à un système à k équations non-linéaires et k inconnues $\widehat{\theta}_1,\ldots,\widehat{\theta}_k$:

$$\text{CN} : \underset{(k\times 1)}{g_n\left(\widehat{\theta},x\right)} = \left.\frac{\partial \ell_n(\theta;x)}{\partial \theta}\right|_{\widehat{\theta}} = \begin{pmatrix} \left.\dfrac{\partial \ell_n(\theta;x)}{\partial \theta_1}\right|_{\widehat{\theta}} \\ \cdots \\ \left.\dfrac{\partial \ell_n(\theta;x)}{\partial \theta_k}\right|_{\widehat{\theta}} \end{pmatrix} = \begin{pmatrix} 0 \\ \ldots \\ 0 \end{pmatrix} \tag{10.83}$$

où $g_n\left(\widehat{\theta},x\right)$ désigne le **vecteur gradient**.

Remarque : Dans de nombreux problèmes, ce système n'admet pas de *solution analytique*. Il convient alors de recourir à une méthode d'optimisation numérique (Gauss-Newton, Newton-Raphson, etc.). On obtient ainsi une solution numérique au problème de maximisation de la log-vraisemblance, c'est-à-dire une réalisation de l'estimateur, sans connaître la forme générale de l'estimateur.

La **matrice hessienne** est une matrice symétrique de dimension $k \times k$ telle que :

$$\underset{(k\times k)}{H_n(\theta,x)} = \frac{\partial^2 \ell_n(\theta;x)}{\partial \theta \partial \theta^\top} = \begin{pmatrix} \dfrac{\partial^2 \ell_n(\theta;x)}{\partial \theta_1^2} & \dfrac{\partial^2 \ell_n(\theta;x)}{\partial \theta_1 \partial \theta_2} & \cdots & \dfrac{\partial^2 \ell_n(\theta;x)}{\partial \theta_1 \partial \theta_k} \\ \dfrac{\partial^2 \ell_n(\theta;x)}{\partial \theta_2 \partial \theta_1} & \dfrac{\partial^2 \ell_n(\theta;x)}{\partial \theta_2^2} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{\partial^2 \ell_n(\theta;x)}{\partial \theta_k \partial \theta_1} & \cdots & \cdots & \dfrac{\partial^2 \ell_n(\theta;x)}{\partial \theta_k^2} \end{pmatrix} \tag{10.84}$$

La condition suffisante du programme de maximisation de la log-vraisemblance est alors la suivante :

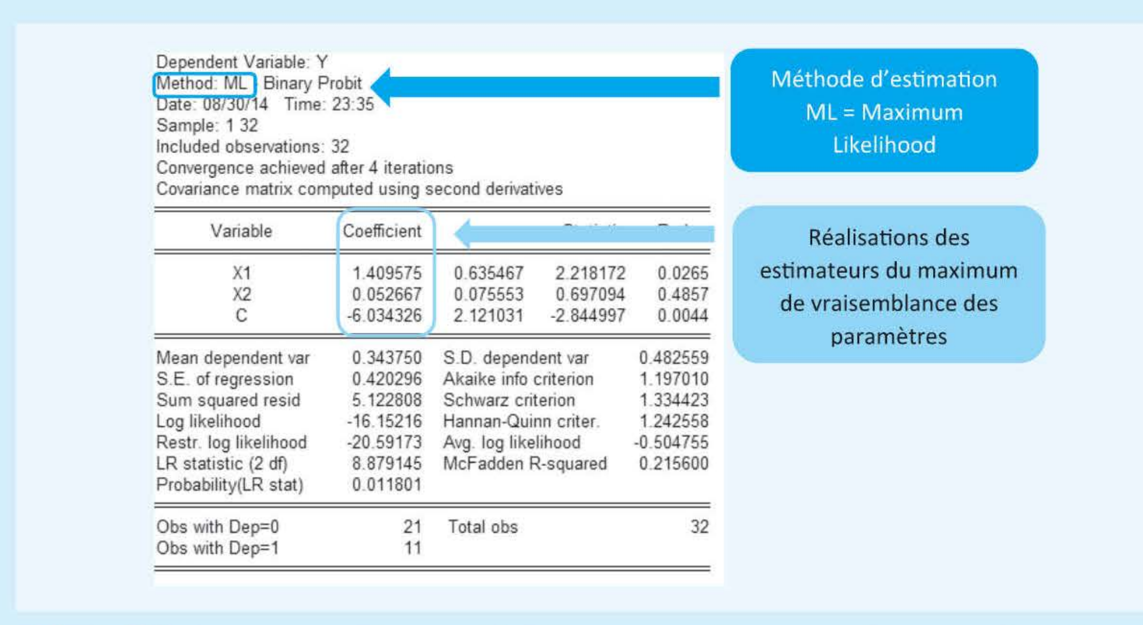
$$\text{CS} : H_n\left(\widehat{\theta},x\right) = \left.\frac{\partial^2 \ell_n(\theta;y|x)}{\partial \theta \partial \theta^\top}\right|_{\widehat{\theta}} \text{ est définie negative} \tag{10.85}$$

Rappelons qu'une matrice est définie négative lorsque toutes ses valeurs propres sont négatives. Considérons un exemple avec deux paramètres ($k = 2$).

EN PRATIQUE

Estimation par maximum de vraisemblance

Dans la pratique, la méthode du maximum de vraisemblance est programmée dans la plupart des logiciels d'économétrie (Eviews, Stata, Rats, SPSS, SAS, Matlab, R, etc.). Comme l'illustre la figure 10.4, lorsque l'on estime les paramètres d'un modèle par maximum de vraisemblance (dans le cas présent un modèle probit), le logiciel affiche les réalisations des estimateurs (estimations ponctuelles). Dans le cas de ce modèle probit, il n'existe pas de solution analytique au programme de maximisation de la log-vraisemblance. Le logiciel utilise alors un algorithme d'optimisation numérique pour déterminer ces estimations. Le message « *convergence achieved after 4 iterations* » indique à l'utilisateur que cet algorithme itératif a convergé vers le maximum de la fonction de log-vraisemblance en 4 itérations.


Dependent Variable: Y
Method: ML Binary Probit
Date: 08/30/14 Time: 23:35
Sample: 1 32
Included observations: 32
Convergence achieved after 4 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient			
X1	1.409575	0.635467	2.218172	0.0265
X2	0.052667	0.075553	0.697094	0.4857
C	-6.034326	2.121031	-2.844997	0.0044

Mean dependent var	0.343750	S.D. dependent var	0.482559
S.E. of regression	0.420296	Akaike info criterion	1.197010
Sum squared resid	5.122808	Schwarz criterion	1.334423
Log likelihood	-16.15216	Hannan-Quinn criter.	1.242558
Restr. log likelihood	-20.59173	Avg. log likelihood	-0.504755
LR statistic (2 df)	8.879145	McFadden R-squared	0.215600
Probability(LR stat)	0.011801		

Obs with Dep=0	21	Total obs	32
Obs with Dep=1	11		

▲ **Figure 10.4 Exemple de sortie du logiciel Eviews : estimation par maximum de vraisemblance d'un modèle probit**

Exemple

On considère un n-échantillon $(Y_1,\ldots,Y_n)$ N.i.d.$\left(m,\sigma^2\right)$ où les paramètres m et σ^2 sont inconnus. On souhaite les estimer par maximum de vraisemblance. Pour cela, on définit un vecteur de paramètres $\theta = \left(m,\sigma^2\right)^{\intercal}$. Pour une réalisation de l'échantillon $(y_1,\ldots,y_n)$, l'estimateur $\widehat{\theta}$ vérifie :

$$\widehat{\theta} = \underset{\sigma^2\in\mathbb{R}^+,m\in\mathbb{R}}{\arg\max}\ \ell_n(\theta;y) \tag{10.86}$$

$$\ell_n(\theta;y) = -\frac{n}{2}\ln\left(\sigma^2\right) - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - m)^2 \tag{10.87}$$

Le gradient de l'échantillon $(y_1,\ldots,y_n)$ est un vecteur 2×1 défini par :

$$\frac{\partial \ell_n(\theta;y)}{\partial \theta} = \begin{pmatrix} \dfrac{\partial \ell_n(\theta;y)}{\partial m} \\ \dfrac{\partial \ell_n(\theta;y)}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(y_i - m) \\ -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4}\displaystyle\sum_{i=1}^{n}(y_i - m)^2 \end{pmatrix} \quad (10.88)$$

La condition nécessaire du programme de maximisation (équations de log-vraisemblance) correspond à un système à deux équations et deux inconnues, à savoir $\widehat{m}$ et $\widehat{\sigma}^2$:

$$g_n\left(\widehat{\theta};y\right) = \left.\frac{\partial \ell_n(\theta;y)}{\partial \theta}\right|_{\widehat{\theta}} = \begin{pmatrix} \dfrac{1}{\widehat{\sigma}^2}\displaystyle\sum_{i=1}^{n}(y_i - \widehat{m}) \\ -\dfrac{n}{2\widehat{\sigma}^2} + \dfrac{1}{2\widehat{\sigma}^4}\displaystyle\sum_{i=1}^{n}(y_i - \widehat{m})^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (10.89)$$

On en déduit une solution :

$$\widehat{\theta} = \begin{pmatrix} \widehat{m} \\ \widehat{\sigma}^2 \end{pmatrix} \quad (10.90)$$

avec

$$\widehat{m} = \frac{1}{n}\sum_{i=1}^{n} y_i = \overline{y}_n, \qquad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y}_n)^2 \quad (10.91)$$

On vérifie que cette solution est un maximum. Pour cela, on construit la matrice hessienne :

$$H_n(\theta;y) = \frac{\partial^2 \ell_n(\theta;y)}{\partial \theta \partial \theta^\top} = \begin{pmatrix} \dfrac{\partial^2 \ell_n(\theta;y)}{\partial m^2} & \dfrac{\partial^2 \ell_n(\theta;y)}{\partial m \partial \sigma^2} \\ \dfrac{\partial^2 \ell_n(\theta;y)}{\partial \sigma^2 \partial m} & \dfrac{\partial^2 \ell_n(\theta;y)}{\partial \sigma^4} \end{pmatrix} \quad (10.92)$$

Dans notre cas, on obtient :

$$H_n(\theta;y) = \begin{pmatrix} -\dfrac{n}{\sigma^2} & -\dfrac{1}{\sigma^4}\displaystyle\sum_{i=1}^{n}(y_i - m) \\ -\dfrac{1}{\sigma^4}\displaystyle\sum_{i=1}^{n}(y_i - m) & \dfrac{n}{2\sigma^4} - \dfrac{1}{\sigma^6}\displaystyle\sum_{i=1}^{n}(y_i - m)^2 \end{pmatrix} \quad (10.93)$$

On évalue la matrice hessienne au point $\widehat{\theta}$:

$$H_n\left(\widehat{\theta};y\right) = \begin{pmatrix} -\dfrac{n}{\widehat{\sigma}^2} & -\dfrac{1}{\widehat{\sigma}^4}\displaystyle\sum_{i=1}^{n}(y_i - \widehat{m}) \\ -\dfrac{1}{\widehat{\sigma}^4}\displaystyle\sum_{i=1}^{n}(y_i - \widehat{m}) & \dfrac{n}{2\widehat{\sigma}^4} - \dfrac{1}{\widehat{\sigma}^6}\displaystyle\sum_{i=1}^{n}(y_i - \widehat{m})^2 \end{pmatrix} \quad (10.94)$$

Puisque $n \times \widehat{m} = \sum_{i=1}^{n} y_i$ et $n \times \widehat{\sigma}^2 = \sum_{i=1}^{n}(y_i - \widehat{m})^2$, on obtient :

$$H_n\left(\widehat{\theta};y\right) = \begin{pmatrix} -\dfrac{n}{\widehat{\sigma}^2} & 0 \\ 0 & \dfrac{n}{2\widehat{\sigma}^4} - \dfrac{n\widehat{\sigma}^2}{\widehat{\sigma}^6} \end{pmatrix} = \begin{pmatrix} -\dfrac{n}{\widehat{\sigma}^2} & 0 \\ 0 & -\dfrac{n}{2\widehat{\sigma}^4} \end{pmatrix} \quad (10.95)$$

Cette matrice diagonale est définie négative car les éléments de sa diagonale sont tous négatifs. Nous avons bien un maximum. Les estimateurs du maximum de vraisemblance des paramètres m et σ^2 sont définis par :

$$\widehat{m} = \overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i, \qquad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \overline{Y}_n\right)^2 \quad (10.96)$$

Notons que l'estimateur du maximum de vraisemblance de la variance σ^2 correspond à la variance empirique non corrigée (▶ chapitre 9).

4 Score, hessienne et quantité d'information de Fisher

Afin d'étudier les propriétés de l'estimateur du maximum de vraisemblance dans la section 5, nous devons à présent définir deux nouveaux concepts : le score et la quantité d'information de Fisher. Nous reviendrons en outre sur la notion de hessienne en introduisant une version stochastique de cette dernière. Dans un premier temps, nous présenterons les définitions dans le cas où le paramètre θ est un scalaire, puis nous étendrons ces définitions au cas vectoriel.

4.1 Score et hessienne

Définition 10.11

Le **score de l'échantillon** $(X_1,\ldots,X_n)$ est une variable aléatoire définie par :

$$S_n(\theta;X) = \frac{\partial \ell_n(\theta;X)}{\partial \theta} \tag{10.97}$$

La forme du score est similaire à celle du gradient. Pourtant, il convient de bien distinguer les deux notions. Rappelons que la fonction de vraisemblance dépend de la réalisation de l'échantillon $(x_1,\ldots,x_n)$. C'est pourquoi la log-vraisemblance est notée sous la forme $\ell_n(\theta;x_1,\ldots,x_n)$, ou de façon plus concise $\ell_n(\theta;x)$, la variable x en minuscule renvoyant à la notion de réalisation. Le gradient est défini comme la dérivée partielle de la fonction de log-vraisemblance par rapport à θ : le **gradient** est donc une **quantité déterministe** (constante), que l'on note $g_n(\theta;x) = \partial \ell_n(\theta;x)/\partial\theta$.

À l'inverse le score correspond à la dérivée d'une « version » stochastique de la fonction de log-vraisemblance dans laquelle on remplace les réalisations $x_1,\ldots,x_n$ par les variables aléatoires de l'échantillon $X_1,\ldots,X_n$: le **score** est donc une **variable aléatoire**. C'est pourquoi on note le score sous la forme $S_n(\theta;X) = \partial \ell_n(\theta;X)/\partial\theta$ où la variable X notée en majuscule renvoie à la notion de variable aléatoire.

Le score étant une variable aléatoire, on peut caractériser sa distribution et ses moments (espérance, variance, etc.). La propriété essentielle du score concerne son espérance.

Propriété

Espérance du score

Pour toute valeur du paramètre $\theta \in \Theta$, le score de l'échantillon vérifie :

$$\mathbb{E}(S_n(\theta;X)) = 0 \tag{10.98}$$

Cette propriété ne s'applique pas au *gradient*. En effet, puisque la quantité $g_n(\theta;x)$ est une constante, son espérance vérifie $\mathbb{E}(g_n(\theta;x)) = g_n(\theta;x)$. Or, cette quantité

n'est pas nulle quelle que soit la valeur du paramètre θ. Le gradient ne s'annule que pour une valeur précise, correspondant à la réalisation de l'estimateur du maximum de vraisemblance :

$$g_n\left(\widehat{\theta}; x\right) = 0 \quad \text{contre} \quad \mathbb{E}\left(S_n(\theta; X)\right) = 0 \;\; \forall\theta \in \Theta \tag{10.99}$$

Remarque : Dans l'énoncé de cette propriété, on précise parfois que l'espérance correspond à l'espérance par rapport à la « vraie » *loi* de la variable X, c'est-à-dire celle obtenue pour la vraie valeur du paramètre θ_0. On note l'espérance par rapport à cette loi sous la forme $\mathbb{E}_{\theta_0}$. Par exemple si X suit une loi exponentielle $\mathcal{E}xp\,(1/\theta)$ et que la vraie valeur du paramètre θ est égale à $\theta_0 = 2$, alors l'espérance $\mathbb{E}_{\theta_0}\left(S_n(\theta; X)\right)$ s'écrit comme :

$$\mathbb{E}_{\theta_0}\left(S_n(\theta; X)\right) = \int_0^{+\infty} S_n(\theta; x) \times f_X(x; \theta_0)\, dx \tag{10.100}$$

$$= \int_0^{+\infty} S_n(\theta; x) \times \frac{1}{2}\exp\left(-\frac{x}{2}\right) dx = 0 \tag{10.101}$$

Le score $S_n(\theta; X)$ est évalué en θ, tandis que la densité est évaluée en θ_0.

Appliquons la propriété du score.

Exemple

Soit un n-échantillon de variables continues, positives et i.i.d. $(D_1,\ldots,D_n)$ admettant une distribution exponentielle $\mathcal{E}xp\,(1/\theta)$ avec $\mathbb{E}(D_i) = \theta$ où $\theta > 0$ est un paramètre inconnu. La fonction de densité de D_i est la suivante :

$$f_{D_i}(d_i; \theta) = \frac{1}{\theta}\exp\left(-\frac{d_i}{\theta}\right), \; \forall d_i \in \mathbb{R}^+ \tag{10.102}$$

Par conséquent la log-vraisemblance de l'échantillon $(d_1,\ldots,d_n)$ est définie par :

$$\ell_n(\theta; d) = -n\ln(\theta) - \frac{1}{\theta}\sum_{i=1}^{n} d_i \tag{10.103}$$

Le *gradient de l'échantillon* (quantité déterministe) est égal à :

$$g_n(\theta; d) = \frac{\partial \ell_n(\theta; d)}{\partial\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} d_i \tag{10.104}$$

La fonction $\ell_n(\theta; D)$ est identique à la log-vraisemblance $\ell_n(\theta; D)$ sauf que les réalisations d_i sont remplacées par les variables aléatoires D_i :

$$\ell_n(\theta; D) = -n\ln(\theta) - \frac{1}{\theta}\sum_{i=1}^{n} D_i \tag{10.105}$$

Le *score de l'échantillon* (variable aléatoire) a une forme similaire au gradient :

$$S_n(\theta; D) = \frac{\partial \ell_n(\theta; D)}{\partial\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} D_i \tag{10.106}$$

On vérifie que son espérance est nulle puisque :

$$\mathbb{E}\left(S_n(\theta; D)\right) = \mathbb{E}\left(-\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} D_i\right) \tag{10.107}$$

$$= -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n}\mathbb{E}(D_i) \tag{10.108}$$

$$= -\frac{n}{\theta} + \frac{n \times \theta}{\theta^2} \tag{10.109}$$

$$= 0 \tag{10.110}$$

De la même façon que l'on distingue le score (aléatoire) et le gradient (déterministe), on peut distinguer deux types d'hessienne : une **hessienne déterministe** (fonction de $x_1,\ldots,x_n$) et une **hessienne stochastique** (fonction des variables aléatoires de l'échantillon $X_1,\ldots,X_n$). Cette dernière est notée de la façon suivante :

$$H_n(\theta; X) = \frac{\partial^2 \ell_n(\theta; X)}{\partial \theta^2} \quad (10.111)$$

Le tableau 10.1 résume les différents concepts de gradient, score et hessienne, ainsi que les notations associées.

▼ **Tableau 10.1 Résumé des différents concepts de score, gradient et hessienne**

Variable aléatoire	Constante
score : $S_n(\theta; X) = \frac{\partial \ell_n(\theta; X)}{\partial \theta}$	gradient : $g_n(\theta; X) = \frac{\partial \ell_n(\theta; x)}{\partial \theta}$
hessiennne : $H_n(\theta; X) = \frac{\partial^2 \ell_n(\theta; X)}{\partial \theta^2}$	hessienne : $H_n(\theta; x) = \frac{\partial^2 \ell_n(\theta; x)}{\partial \theta^2}$

4.2 Information de Fisher

À partir de ces éléments, nous pouvons à présent définir la quantité d'**information de Fisher**, du nom du statisticien britannique Ronald A. Fisher (1890-1962). Cette quantité est notamment utilisée pour montrer qu'un estimateur est **efficace** au sens de la borne FDCR ou de Cramer-Rao (▶ chapitre 9).

Définition 10.12

La quantité d'**information de Fisher** associée à l'**échantillon** est une constante définie par la variance du score ou l'espérance de l'opposée de la hessienne stochastique :

$$\mathrm{I}_n(\theta) = \mathbb{V}(S_n(\theta; X)) = \mathbb{E}(-H_n(\theta; X)) \quad (10.112)$$

Exemple

Soit un n-échantillon de variables positives $(D_1,\ldots,D_n)$ i.i.d. admettant une distribution exponentielle $\mathcal{E}xp(1/\theta)$ où $\theta > 0$ est un paramètre inconnu. D'après les propriétés de la loi exponentielle, nous savons que $\mathbb{E}(D_i) = \theta$ et $\mathbb{V}(D_i) = \theta^2$. La densité de la loi exponentielle est définie par :

$$f_D(d; \theta) = \frac{1}{\theta} \exp\left(-\frac{d}{\theta}\right), \quad \forall d \in \mathbb{R}^+ \quad (10.113)$$

On en déduit la log-vraisemblance de $(d_1,\ldots,d_n)$, le score et la hessienne (stochastique) :

$$\ell_n(\theta; d) = -n \ln(\theta) - \frac{1}{\theta} \sum_{i=1}^{n} d_i \quad (10.114)$$

$$S_n(\theta; D) = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^{n} D_i \quad (10.115)$$

$$H_n(\theta; D_n) = \frac{n}{\theta^2} - \frac{2}{\theta^3} \sum_{i=1}^{n} D_i \quad (10.116)$$

Déterminons à présent la quantité d'*information de Fisher associée à l'échantillon*. Utilisons pour cela la première définition (variance du score) :

$$\mathrm{I}_n(\theta) = \mathbb{V}(S_n(\theta; D)) = \mathbb{V}\left(-\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} D_i\right) \tag{10.117}$$

$$= \frac{1}{\theta^4}\sum_{i=1}^{n}\mathbb{V}(D_i) = \frac{n \times \theta^2}{\theta^4} = \frac{n}{\theta^2} \tag{10.118}$$

On vérifie qu'en utilisant la seconde définition (opposée de l'espérance de la hessienne), on obtient la même quantité :

$$\mathrm{I}_n(\theta) = \mathbb{E}(-H_n(\theta; X)) = \mathbb{E}\left(-\frac{n}{\theta^2} + \frac{2}{\theta^3}\sum_{i=1}^{n} D_i\right) \tag{10.119}$$

$$= -\frac{n}{\theta^2} + \frac{2}{\theta^3}\sum_{i=1}^{n}\mathbb{E}(D_i) = -\frac{n}{\theta^2} + \frac{2 \times n \times \theta}{\theta^3} = \frac{n}{\theta^2} \tag{10.120}$$

Remarque : Le score ayant une espérance nulle, il existe une troisième formule équivalente pour définir la quantité d'information de Fisher de l'échantillon. En effet :

$$\mathbb{V}(S_n(\theta; X)) = \mathbb{E}(S_n^2(\theta; X)) - (\mathbb{E}(S_n(\theta; X)))^2 = \mathbb{E}(S_n^2(\theta; X)) \tag{10.121}$$

Par conséquent, la quantité d'information de Fisher peut être définie par :

$$\mathrm{I}_n(\theta) = \mathbb{E}\left(S_n^2(\theta; X)\right) \tag{10.122}$$

Finalement, on peut étendre ces définitions (score, hessienne, information de Fisher) non plus au cas d'un **échantillon**, mais au cas d'une **observation** x_i particulière. De façon générale, si l'on note $\mathrm{I}_i(\theta)$ la quantité d'information de Fisher associée à la $i^{ème}$ observation de l'échantillon, on a bien sûr une relation du type :

$$\mathrm{I}_n(\theta) = \sum_{i=1}^{n}\mathrm{I}_i(\theta) \tag{10.123}$$

avec par définition :

$$\mathrm{I}_i(\theta) = \mathbb{V}(S_i(\theta; X_i)) = \mathbb{E}(-H_i(\theta; X_i)) \tag{10.124}$$

où $S_i(\theta; X_i)$ et $H_i(\theta; X_i)$ désignent respectivement le score et la hessienne associés à la variable X_i pour $i \in \{1,\ldots,n\}$.

$$S_i(\theta; X_i) = \frac{\partial \ell_i(\theta; X_i)}{\partial \theta} \qquad H_i(\theta; X_i) = \frac{\partial^2 \ell_i(\theta; X_i)}{\partial \theta^2} \tag{10.125}$$

Lorsque l'on considère des **distributions marginales** (c'est-à-dire lorsque l'on ne considère pas un modèle économétrique), les quantités d'information de Fisher de toutes les observations $i = 1,\ldots,n$ sont strictement identiques : la quantité $\mathrm{I}_i(\theta)$ ne dépend pas de l'indice i.

$$\mathrm{I}_i(\theta) = \mathrm{I}(\theta) \tag{10.126}$$

D'après l'équation (10.123), la somme des quantités de Fisher individuelles correspond à celle de l'échantillon. On obtient ainsi une relation évidente entre les quantités $\mathrm{I}(\theta)$ et $\mathrm{I}_n(\theta)$:

$$I_n(\theta) = \sum_{i=1}^{n} I_i(\theta) = n \times I(\theta) \tag{10.127}$$

Dans le cas d'un modèle économétrique, la quantité d'information de Fisher associée à une observation et la quantité moyenne diffèrent, mais on peut toujours définir cette dernière de la façon suivante.

Définition 10.13

On appelle **quantité moyenne** d'information de Fisher, la quantité $\mathrm{I}(\theta)$ telle que :

$$\mathrm{I}(\theta) = \frac{1}{n} \times \mathrm{I}_n(\theta) \tag{10.128}$$

Nous utiliserons cette quantité moyenne pour caractériser la distribution asymptotique de l'estimateur du maximum de vraisemblance dans la section 5.

Exemple

Reprenons l'exemple précédent d'un échantillon de variables $D_1,\ldots,D_n$ i.i.d. suivant une loi exponentielle $\mathcal{E}xp(1/\theta)$ avec $\theta > 0$, $\mathbb{E}(D_i) = \theta$ et $\mathbb{V}(D_i) = \theta^2$, $\forall i = 1,\ldots,n$. La log-vraisemblance associée à une observation d_i correspond au logarithme de la densité :

$$\ell_i(\theta; d_i) = \ln f_{D_i}(d_i; \theta) = -\ln(\theta) - \frac{d_i}{\theta} \tag{10.129}$$

Le score et la hessienne associés à la variable D_i sont définis par :

$$S_i(\theta; D_i) = \frac{\partial \ell_i(\theta; D_i)}{\partial \theta} = -\frac{1}{\theta} + \frac{D_i}{\theta^2} \tag{10.130}$$

$$H_i(\theta; D_i) = \frac{\partial^2 \ell_i(\theta; D_i)}{\partial \theta^2} = \frac{1}{\theta^2} - \frac{2D_i}{\theta^3} \tag{10.131}$$

Selon la première définition (variance du score), la quantité d'information de Fisher associée à cette observation est égale à :

$$\mathrm{I}_i(\theta) = \mathbb{V}(S_i(\theta; D_i)) = \mathbb{V}\left(-\frac{1}{\theta} + \frac{D_i}{\theta^2}\right) = \frac{1}{\theta^4}\mathbb{V}(D_i) = \frac{1}{\theta^2} \tag{10.132}$$

On vérifie que la seconde définition (opposée de l'espérance de la hessienne) donne la même quantité :

$$\mathrm{I}_i(\theta) = \mathbb{E}(-H_i(\theta; D_i)) = \mathbb{E}\left(-\left(\frac{1}{\theta^2} - \frac{2D_i}{\theta^3}\right)\right) \tag{10.133}$$

$$= -\frac{1}{\theta^2} + \frac{2 \times \mathbb{E}(D_i)}{\theta^3} = -\frac{1}{\theta^2} + \frac{2 \times \theta}{\theta^3} = \frac{1}{\theta^2} \tag{10.134}$$

Dans ce cas, la quantité d'information associée à la variable X_i ne dépend pas de l'indice i : elle correspond à la quantité moyenne d'information de Fisher.

$$\mathrm{I}_i(\theta) = \mathrm{I}(\theta) = \frac{1}{\theta^2} \tag{10.135}$$

Dans l'exemple précédent, nous avions vu que la quantité d'information de Fisher associée à l'*échantillon* était égale à :

$$\mathrm{I}_n(\theta) = \frac{n}{\theta^2} \tag{10.136}$$

On a donc une relation entre la quantité d'information de Fisher associée à l'*échantillon* et la quantité moyenne d'information de Fisher du type :

$$\mathrm{I}_n(\theta) = n \times \mathrm{I}(\theta) \tag{10.137}$$

FOCUS

Modèle économétrique et information de Fisher

Dans le cadre d'un modèle économétrique, la log-vraisemblance est fondée sur la distribution conditionnelle de la variable endogène Y sachant que la variable explicative est fixée à une certaine valeur, *i.e.* $X = x$. Soit $\ell_n(\theta; Y|x)$ la log-vraisemblance conditionnelle de l'échantillon $(y_i,x_i)_{i=1}^n$. On peut définir le score et la hessienne associés à une variable Y_i conditionnellement à $X_i = x_i$ de la façon usuelle :

$$S_i(\theta; Y_i|x_i) = \frac{\partial \ell_i(\theta; Y_i|x_i)}{\partial \theta}, \qquad (10.138)$$

$$H_i(\theta; Y_i|x_i) = \frac{\partial^2 \ell_i(\theta; Y_i|x_i)}{\partial \theta^2} \qquad (10.139)$$

La quantité d'information associée à cette variable Y_i est définie par la variance du score ou l'espérance de l'opposée de la hessienne :

$$\mathrm{I}_i(\theta) = \mathbb{V}(S_i(\theta; Y_i|x_i)) = \mathbb{E}(-H_i(\theta; Y_i|x_i)) \qquad (10.140)$$

Mais cette quantité dépend alors de l'observation x_i. Par conséquent, la quantité d'information de Fisher associée au $i^{ème}$ individu de l'échantillon peut ne pas être identique à celle du $j^{ème}$ individu :

$$\mathrm{I}_i(\theta) \neq \mathrm{I}_j(\theta) \quad \text{pour } i \neq j \quad \text{si } x_i \neq x_j \qquad (10.141)$$

Dans ce cas, la quantité moyenne d'information de Fisher est définie par l'espérance de la quantité individuelle associée à la variable Y_i :

$$\mathrm{I}(\theta) = \mathbb{E}_X(\mathrm{I}_i(\theta)) \qquad (10.142)$$

où le terme $\mathbb{E}_X$ désigne l'espérance par rapport à la distribution de la variable explicative X. L'idée est de construire une sorte de « moyenne » des quantités individuelles pour toutes les valeurs possibles de X. Par construction, on retrouve alors l'égalité :

$$\mathrm{I}(\theta) = \frac{1}{n} \times \mathrm{I}_n(\theta) \qquad (10.143)$$

Ainsi, dans le cas d'un modèle économétrique, il devient important d'indicer les espérances afin d'éviter les confusions. La définition de la quantité moyenne d'information de Fisher devient :

$$\begin{aligned}\mathrm{I}(\theta) &= \mathbb{E}_X(\mathbb{V}_{\theta_0}(S_i(\theta; Y_i|x_i))) \\ &= \mathbb{E}_X(\mathbb{E}_{\theta_0}(-H_i(\theta; Y_i|x_i)))\end{aligned} \qquad (10.144)$$

où $\mathbb{E}_X$ désigne l'espérance par rapport à la distribution de la variable explicative X et $\mathbb{E}_{\theta_0}$ désigne l'espérance par rapport à la vraie loi conditionnelle de Y sachant $X = x$.

4.3 Extension au cas avec plusieurs paramètres

Nous pouvons à présent étendre les définitions précédentes (score, hessienne, information de Fisher) au cas où $\theta = (\theta_1,...,\theta_k)^\top$ désigne un vecteur de k paramètres. Le **score** correspond alors à un vecteur de dimension $k \times 1$ tel que :

$$\underset{(k\times 1)}{S_n(\theta; X)} = \frac{\partial \ell_n(\theta; X)}{\partial \theta} = \begin{pmatrix} \dfrac{\partial \ell_n(\theta; X)}{\partial \theta_1} \\ \dots \\ \dfrac{\partial \ell_n(\theta; X)}{\partial \theta_k} \end{pmatrix} \qquad (10.145)$$

Par définition, le score vérifie :

$$\mathbb{E}(\underset{(k\times 1)}{S_n(\theta; X)}) = \underset{(k\times 1)}{0} \qquad (10.146)$$

La **matrice hessienne** (stochastique) est une matrice de dimension $k \times k$ similaire à celle de l'équation (10.84) :

$$\underset{(k\times k)}{H_n(\theta,X)} = \frac{\partial^2 \ell_n(\theta;X)}{\partial\theta\partial\theta^\top} = \begin{pmatrix} \frac{\partial^2 \ell_n(\theta;X)}{\partial\theta_1^2} & \frac{\partial^2 \ell_n(\theta;X)}{\partial\theta_1\partial\theta_2} & \dots & \frac{\partial^2 \ell_n(\theta;X)}{\partial\theta_1\partial\theta_k} \\ \frac{\partial^2 \ell_n(\theta;X)}{\partial\theta_2\partial\theta_1} & \frac{\partial^2 \ell_n(\theta;X)}{\partial\theta_2^2} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 \ell_n(\theta;X)}{\partial\theta_k\partial\theta_1} & \dots & \dots & \frac{\partial^2 \ell_n(\theta;X)}{\partial\theta_k^2} \end{pmatrix} \quad (10.147)$$

La **matrice d'information de Fisher** de l'échantillon est une matrice de dimension $k \times k$ définie par l'une des trois relations suivantes :

$$\mathrm{I}_n(\theta) = \mathbb{V}(S_n(\theta;X)) = \mathbb{E}(-H_n(\theta;X)) = \mathbb{E}\left(S_n(\theta;X) \times S_n(\theta;X)^\top\right) \quad (10.148)$$

Reprenons l'exemple précédent d'un échantillon normal.

Exemple

On considère un n-échantillon $(Y_1,\dots,Y_n)$ N.i.d.$\left(m,\sigma^2\right)$ où les paramètres m et σ^2 sont inconnus. On définit un vecteur de paramètres $\theta = \left(m,\sigma^2\right)^\top$ avec $k = 2$. La log-vraisemblance de l'échantillon $(y_1,\dots,y_n)$ est définie par :

$$\ell_n(\theta;y) = -\frac{n}{2}\ln\left(\sigma^2\right) - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - m)^2 \quad (10.149)$$

Le *score* de l'échantillon est un vecteur 2×1 :

$$S_n(\theta;Y) = \frac{\partial \ell_n(\theta;Y)}{\partial\theta} = \begin{pmatrix} \frac{\partial \ell_n(\theta;Y)}{\partial m} \\ \frac{\partial \ell_n(\theta;Y)}{\partial\sigma^2} \end{pmatrix} \quad (10.150)$$

$$= \begin{pmatrix} \frac{1}{\sigma^2}\sum_{i=1}^n (Y_i - m) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (Y_i - m)^2 \end{pmatrix} \quad (10.151)$$

La *matrice hessienne* (stochastique) est une matrice 2×2 définie par :

$$H_n(\theta;Y) = \frac{\partial^2 \ell_n(\theta;Y)}{\partial\theta\partial\theta^\top} = \begin{pmatrix} \frac{\partial^2 \ell_n(\theta;Y)}{\partial m^2} & \frac{\partial^2 \ell_n(\theta;Y)}{\partial m\partial\sigma^2} \\ \frac{\partial^2 \ell_n(\theta;Y)}{\partial\sigma^2\partial m} & \frac{\partial^2 \ell_n(\theta;Y)}{\partial\sigma^4} \end{pmatrix} \quad (10.152)$$

$$= \begin{pmatrix} -\frac{n}{\sigma^2} & -\frac{1}{\sigma^4}\sum_{i=1}^n (Y_i - m) \\ -\frac{1}{\sigma^4}\sum_{i=1}^n (Y_i - m) & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^n (Y_i - m)^2 \end{pmatrix} \quad (10.153)$$

Par conséquent, la matrice d'information de Fisher associée à l'échantillon est égale à l'espérance de l'opposée de la hessienne (première définition) :

$$\mathrm{I}_n(\theta) = \mathbb{E}\left(-H_n(\theta; Y)\right) \quad (10.154)$$

$$= \mathbb{E}\begin{pmatrix} \dfrac{n}{\sigma^2} & \dfrac{1}{\sigma^4}\displaystyle\sum_{i=1}^{n}(Y_i - m) \\ \dfrac{1}{\sigma^4}\displaystyle\sum_{i=1}^{n}(Y_i - m) & -\dfrac{n}{2\sigma^4} + \dfrac{1}{\sigma^6}\displaystyle\sum_{i=1}^{n}(Y_i - m)^2 \end{pmatrix} \quad (10.155)$$

$$= \begin{pmatrix} \dfrac{n}{\sigma^2} & \dfrac{1}{\sigma^4}\displaystyle\sum_{i=1}^{n}\mathbb{E}(Y_i - m) \\ \dfrac{1}{\sigma^4}\displaystyle\sum_{i=1}^{n}\mathbb{E}(Y_i - m) & -\dfrac{n}{2\sigma^4} + \dfrac{1}{\sigma^6}\displaystyle\sum_{i=1}^{n}\mathbb{E}\left((Y_i - m)^2\right) \end{pmatrix} \quad (10.156)$$

Puisque $\mathbb{E}(Y_i) = m$, on a $\mathbb{E}(Y_i - m) = 0$. De plus, par définition de la variance $\mathbb{E}((Y_i - m)^2) = \sigma^2$. On montre ainsi que la matrice d'information de Fisher associée à l'échantillon est une matrice 2×2 symétrique et définie positive telle que :

$$\mathrm{I}_n(\theta) = \begin{pmatrix} \dfrac{n}{\sigma^2} & 0 \\ 0 & \dfrac{n}{2\sigma^4} \end{pmatrix} \quad (10.157)$$

La matrice d'information moyenne de Fisher est alors égale à :

$$\mathrm{I}(\theta) = \frac{1}{n} \times \mathrm{I}_n(\theta) = \begin{pmatrix} \dfrac{1}{\sigma^2} & 0 \\ 0 & \dfrac{1}{2\sigma^4} \end{pmatrix} \quad (10.158)$$

5 Propriétés du maximum de vraisemblance

La question qui se pose à présent est de savoir si l'estimateur du maximum de vraisemblance est un « bon » estimateur. Est-il sans biais, efficace et convergent ? Quelle est sa distribution asymptotique ? Afin d'étudier ces **propriétés** nous allons poser des hypothèses sur la distribution de la variable d'intérêt X. Ces hypothèses sont qualifiées d'hypothèses de régularité.

Les **hypothèses de régularité** sont au nombre de trois :

- **Hypothèse 1 :** la fonction $\ln f_X(\theta; x_i)$ est trois fois différentiable par rapport à θ. Ses dérivées sont **continues** et **finies** pour toute valeur de x et de θ.
- **Hypothèse 2 :** Les **espérances** des dérivées première et seconde de $\ln f_X(\theta; X_i)$ par rapport à θ existent.
- **Hypothèse 3 :** la vraie valeur de θ, notée θ_0, appartient à un ensemble compact Θ.

Sous ces *hypothèses de régularité*, on peut montrer que l'estimateur du maximum de vraisemblance présente de bonnes propriétés :

1. L'estimateur du maximum de vraisemblance est **convergent**.
2. L'estimateur du maximum de vraisemblance est asymptotiquement **efficace**.
3. L'estimateur du maximum de vraisemblance est asymptotiquement **normalement distribué**.

Propriété

Convergence

Sous les hypothèses de régularité, l'estimateur du maximum de vraisemblance $\widehat{\theta}$ est **convergent** (au sens faible) :

$$\widehat{\theta} \xrightarrow{p} \theta_0 \qquad (10.159)$$

où θ_0 désigne la vraie valeur du paramètre θ.

Cette propriété[4] est particulièrement importante car elle implique que les réalisations de l'estimateur $\widehat{\theta}$ auront de grandes chances d'être très concentrées autour de la vraie valeur du paramètre si la taille d'échantillon n est suffisamment grande (▶ chapitre 9). Considérons un exemple.

Exemple

Soit un n-échantillon $(D_1,\ldots,D_n)$ de variables positives et i.i.d. admettant une distribution exponentielle $\mathcal{E}xp\,(1/\theta)$. Le logarithme de la densité de la loi exponentielle, défini par :

$$\ln f_D\,(d;\theta) = -\ln(\theta) - \frac{d}{\theta} \quad \forall d \in \mathbb{R}^+ \qquad (10.160)$$

vérifie les hypothèses de régularité. La log-vraisemblance de l'échantillon $(d_1,\ldots,d_n)$ est égale à :

$$\ell_n\,(\theta; d) = -n\ln(\theta) - \frac{1}{\theta}\sum_{i=1}^{n} d_i \qquad (10.161)$$

On admet que l'estimateur du maximum de vraisemblance est défini par :

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} D_i \qquad (10.162)$$

Montrons que cet estimateur est convergent. Soit θ_0 la vraie valeur du paramètre. D'après les propriétés de la loi exponentielle, nous savons que $\mathbb{E}\,(D_i) = \theta_0$ et $\mathbb{V}\,(D_i) = \theta_0^2$. Dès lors, il vient :

$$\mathbb{E}\left(\widehat{\theta}\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} D_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\,(D_i) = \frac{n \times \theta_0}{n} = \theta_0 \qquad (10.163)$$

L'estimateur $\widehat{\theta}$ est sans biais. De plus :

$$\mathbb{V}\left(\widehat{\theta}\right) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} D_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}\,(D_i) = \frac{n \times \theta_0^2}{n^2} = \frac{\theta_0^2}{n} \qquad (10.164)$$

Par conséquent :

$$\mathbb{E}\left(\widehat{\theta}\right) = \theta_0 \qquad \lim_{n\to\infty}\mathbb{V}\left(\widehat{\theta}\right) = 0 \qquad (10.165)$$

Donc l'estimateur $\widehat{\theta}$ est convergent (au sens faible) :

$$\widehat{\theta} \xrightarrow{p} \theta_0 \qquad (10.166)$$

4 Pour une démonstration de cette propriété, voir Amemiya (1985).

Sous des hypothèses de régularité plus strictes, il est possible de montrer que l'estimateur du maximum de vraisemblance $\widehat{\theta}$ est convergent au sens fort, c'est-à-dire qu'il converge presque sûrement vers θ_0.

Propriété

Efficacité

Sous les hypothèses de régularité, l'estimateur du maximum de vraisemblance est **efficace**. Sa variance atteint la borne FDCR ou borne de Cramer-Rao :

$$\mathbb{V}\left(\widehat{\theta}\right) = \mathrm{I}_n^{-1}(\theta_0) \tag{10.167}$$

où $\mathrm{I}_n(\theta_0)$ désigne la quantité d'information de Fisher associée à l'échantillon et évaluée au point θ_0, vraie valeur du paramètre.

Ainsi, sous les hypothèses de régularité, l'estimateur du maximum de vraisemblance a la plus faible variance possible comparativement à celles de tous les estimateurs sans biais. C'est donc un estimateur relativement précis. Considérons un exemple.

Exemple

Soit un n-échantillon $(X_1,\ldots,X_n)$ de variables aléatoires discrètes i.i.d., définies sur $\mathbb{N}$ et admettant une distribution de Poisson de paramètre $\theta > 0$ telle que :

$$f_{X_i}(x_i;\theta) = \Pr(X_i = x_i) = \exp(-\theta)\frac{\theta^{x_i}}{x_i!} \quad \forall x_i \in \mathbb{N} \tag{10.168}$$

La fonction $\ln f_{X_i}(x_i;\theta)$ satisfait les hypothèses de régularité. Montrons que l'estimateur du maximum de vraisemblance $\widehat{\theta}$ du paramètre θ est efficace au sens de la borne FDCR. La log-vraisemblance de l'échantillon $(x_1,\ldots,x_n)$ est égale à :

$$\ell_n(\theta;x) = -\theta + \ln(\theta)\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\ln(x_i!) \tag{10.169}$$

On admet que l'estimateur du maximum de vraisemblance est défini par :

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{10.170}$$

Sachant que $\mathbb{E}(X_i) = \mathbb{V}(X_i) = \theta_0$, où θ_0 désigne la vraie valeur du paramètre θ, on en déduit la variance de l'estimateur $\widehat{\theta}$:

$$\mathbb{V}\left(\widehat{\theta}\right) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i) = \frac{n\times\theta_0}{n^2} = \frac{\theta_0}{n} \tag{10.171}$$

Déterminons la borne FDCR. Le score et la hessienne (stochastique) associés à l'échantillon sont respectivement définis par :

$$S_n(\theta;X) = \frac{\partial \ln L_n(\theta;X)}{\partial\theta} = -n + \frac{1}{\theta}\sum_{i=1}^{n} X_i \tag{10.172}$$

$$H_n(\theta;X) = \frac{\partial^2 \ln L_n(\theta;X)}{\partial\theta^2} = -\frac{1}{\theta^2}\sum_{i=1}^{n} X_i \tag{10.173}$$

La quantité d'information de Fisher associée à l'échantillon est égale à l'espérance de l'opposée de la hessienne :

$$\mathrm{I}_n(\theta_0) = \mathbb{E}(-H_n(\theta_0;X)) = \mathbb{E}\left(\frac{1}{\theta_0^2}\sum_{i=1}^{n} X_i\right) \tag{10.174}$$

$$= \frac{1}{\theta_0^2}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{n\times\theta_0}{\theta_0^2} = \frac{n}{\theta_0} \tag{10.175}$$

On vérifie que la variance de $\widehat{\theta}$ atteint la borne FDCR :

$$\mathbb{V}\left(\widehat{\theta}\right) = \mathrm{I}_n^{-1}(\theta_0) = \frac{\theta_0}{n} \tag{10.176}$$

Par conséquent, l'estimateur du maximum de vraisemblance $\widehat{\theta}$ est efficace.

Propriété

Distribution asymptotique

Sous les hypothèses de régularité, l'estimateur du maximum de vraisemblance $\widehat{\theta}$ est **asymptotiquement normalement distribué** :

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \overset{d}{\rightarrow} \mathcal{N}\left(0, \mathrm{I}^{-1}(\theta_0)\right) \tag{10.177}$$

où θ_0 désigne la vraie valeur du paramètre et $\mathrm{I}(\theta_0)$ correspond à la quantité d'information moyenne de Fisher évaluée au point θ_0.

Comme nous l'avons vu dans le chapitre 9, une autre façon de comprendre ce résultat est la suivante.

Corollaire 10.1

Distribution asymptotique

Pour une taille d'échantillon n suffisamment importante, l'estimateur du maximum de vraisemblance $\widehat{\theta}$ est asymptotiquement et approximativement distribué selon une loi normale :

$$\widehat{\theta} \overset{asy}{\approx} \mathcal{N}\left(\theta_0, \frac{1}{n} \times \mathrm{I}^{-1}(\theta_0)\right) \tag{10.178}$$

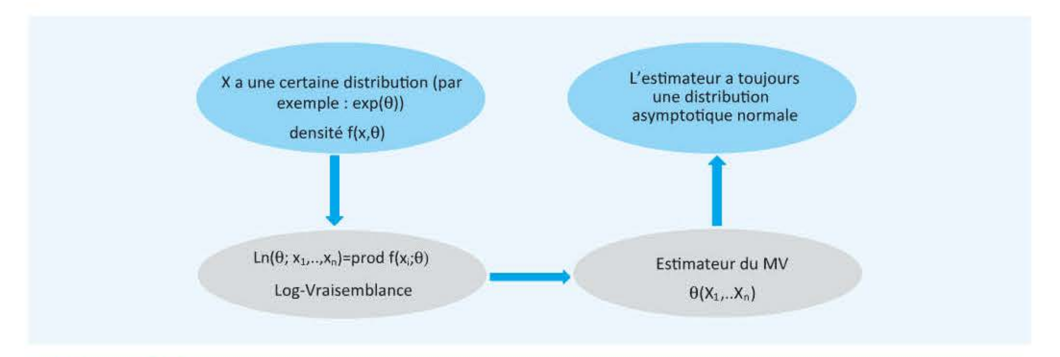
Puisque $\mathrm{I}_n(\theta_0) = n \times \mathrm{I}(\theta_0)$, ce résultat peut se réécrire sous la forme suivante :

$$\widehat{\theta} \overset{asy}{\approx} \mathcal{N}\left(\theta_0, \mathrm{I}_n^{-1}(\theta_0)\right) \tag{10.179}$$

où $\mathrm{I}_n(\theta_0)$ désigne la quantité d'information de Fisher associée à l'échantillon, évaluée au point θ_0.

Le caractère général, certains diront « magique », de la méthode du maximum de vraisemblance réside principalement dans ce résultat : quel que soit le problème posé, s'il est régulier, la distribution asymptotique de l'estimateur du maximum de vraisemblance est toujours normale. Comme l'illustre la figure 10.5, on considère un échantillon de variables ayant une certaine distribution (Poisson, exponentielle, Student, khi-deux, etc.) et l'on construit une fonction, à savoir la fonction de vraisemblance. En maximisant cette fonction, on obtient la forme d'un estimateur qui est une variable aléatoire. Or, quel que soit le problème initial, cet estimateur a toujours une distribution asymptotique normale. Cette « magie » ne tient en fait qu'à l'application du théorème central limite au score de l'échantillon (pour une démonstration formelle, voir Amemiya, 1985).

La variance asymptotique de l'estimateur du maximum de vraisemblance correspond à la variance de sa distribution asymptotique.



▲ **Figure 10.5** **Caractère général de la méthode du maximum de vraisemblance**

Définition 10.14

La **variance asymptotique** de l'estimateur du maximum de vraisemblance $\widehat{\theta}$ est égale à :

$$\mathbb{V}_{asy}\left(\widehat{\theta}\right) = \mathrm{I}_n^{-1}\left(\theta_0\right)$$

où $\mathrm{I}_n\left(\theta_0\right)$ désigne la quantité d'information de Fisher associée à l'échantillon, évaluée au point θ_0. Cette variance asymptotique correspond à la borne FDCR.

Ce résultat confirme le caractère efficace (au sens de la borne FDCR) de l'estimateur du maximum de vraisemblance.

EN PRATIQUE

Estimation de la matrice de variance-covariance asymptotique

De très nombreux logiciels d'économétrie permettent d'estimer les paramètres de modèles par la méthode du maximum de vraisemblance. Tous ces logiciels reportent (i) les estimations des paramètres, c'est-à-dire les réalisations $\widehat{\theta}(x)$ et (ii) les écarts-types associés à ces estimateurs (*standard errors*). Plus précisément, il s'agit de la réalisation des estimateurs des écarts-types (racines carrées des variances) asymptotiques des estimateurs $\widehat{\theta}$. Ces écarts-types sont donc construits à partir d'un estimateur de la matrice de variance-covariance asymptotique de l'estimateur du maximum de vraisemblance $\widehat{\theta}$. Comment **estimer** la **matrice de variance-covariance asymptotique** $\mathbb{V}_{asy}(\widehat{\theta})$? Nous savons que cette matrice correspond à l'inverse de la matrice d'information de Fisher associée à l'échantillon et évaluée au point θ_0 :

$$\mathbb{V}_{asy}(\widehat{\theta}) = \mathrm{I}_n^{-1}\left(\theta_0\right) \quad (10.180)$$

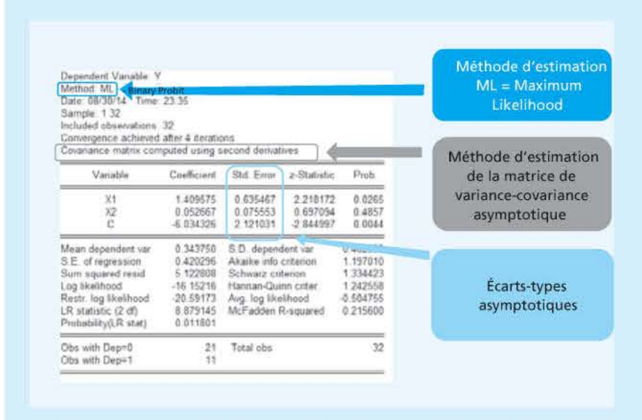
Bien évidemment, puisque θ_0 est inconnu, on ne connaît pas la matrice $\mathrm{I}(\theta_0)$: il convient de l'estimer. Si $\widehat{\theta}$ converge en probabilité vers θ_0, alors les trois estimateurs suivants :

$$\widehat{\mathrm{I}}_n\left(\widehat{\theta}\right) = \sum_{i=1}^{n} \widehat{\mathrm{I}}_i\left(\widehat{\theta}\right) \quad (10.181)$$

$$\widehat{\mathrm{I}}_n\left(\widehat{\theta}\right) = \sum_{i=1}^{n} \left(\left.\frac{\partial \ell_i\left(\theta; x_i\right)}{\partial \theta}\right|_{\widehat{\theta}} \left.\frac{\partial \ell_i\left(\theta; x_i\right)}{\partial \theta}\right|_{\widehat{\theta}}^{\top} \right) \quad (10.182)$$

$$\widehat{\mathrm{I}}_n\left(\widehat{\theta}\right) = \sum_{i=1}^{n} \left(- \left.\frac{\partial^2 \ell_i\left(\theta; x_i\right)}{\partial \theta \partial \theta^{\top}}\right|_{\widehat{\theta}} \right) \quad (10.183)$$

sont des estimateurs convergents de la matrice $\mathrm{I}_n(\theta_0)$. Le premier estimateur (équation (10.181)) correspond à la moyenne des n matrices d'information de Fisher individuelles (pour $x_1,\ldots,x_n$) évaluées au point $\widehat{\theta}$. Cet estimateur est rarement disponible en pratique. Le second estimateur (équation (10.182)) correspond à la moyenne des produits des gradients individuels évalués au point $\widehat{\theta}$. Il est connu sous le nom d'**estimateur BHHH** pour Berndt, Hall, Hall, et Hausman. Le troisième estimateur (équation (10.183)) correspond à l'opposée de la moyenne des hessiennes individuelles. C'est l'estimateur le plus utilisé dans les logiciels d'économétrie.



▲ **Figure 10.6 Exemple de sortie du logiciel Eviews : estimation par maximum de vraisemblance d'un modèle probit**

L'illustration de la figure 10.6 reproduit un exemple de sortie du logiciel Eviews obtenue pour une estimation par maximum de vraisemblance des paramètres d'un modèle probit (modèle dichotomique utilisé notamment pour les procédures de scoring en marketing ou dans le domaine du risque bancaire). Dans cette sortie, la phrase « *Covariance matrix computed using second derivatives* » signifie que la matrice de variance-covariance asymptotique a été estimée à partir de la hessienne, c'est-à-dire à partir des dérivées secondes de la log-vraisemblance (équation (10.183)).

Une fois que l'on a estimé la matrice d'information de Fisher, il suffit de l'inverser pour obtenir un estimateur de la matrice de variance-covariance asymptotique :

$$\widehat{\mathbb{V}}_{asy}(\widehat{\theta}) = \widehat{\mathrm{I}}_n^{-1}(\theta_0) \quad (10.184)$$

Les écarts-types correspondent alors à la racine carrée des éléments de la diagonale de cette matrice :

$$\widehat{\mathbb{V}}_{asy}(\widehat{\theta}) = \begin{pmatrix} \boxed{\widehat{\mathbb{V}}_{asy}(\widehat{\theta}_1)} & \widehat{\mathbb{C}ov}_{asy}(\widehat{\theta}_1,\widehat{\theta}_2) \ldots \widehat{\mathbb{C}ov}_{asy}(\widehat{\theta}_1,\widehat{\theta}_k) & & \\ \widehat{\mathbb{C}ov}_{asy}(\widehat{\theta}_2,\widehat{\theta}_1) & \boxed{\widehat{\mathbb{V}}_{asy}(\widehat{\theta}_2)} & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \widehat{\mathbb{C}ov}_{asy}(\widehat{\theta}_k,\widehat{\theta}_1) & \ldots & \ldots & \boxed{\widehat{\mathbb{V}}_{asy}(\widehat{\theta}_k)} \end{pmatrix} \quad (10.185)$$

Exemple

Soit un n-échantillon de variables continues, positives et i.i.d. $(D_1,\ldots,D_n)$ admettant une distribution exponentielle $\mathcal{E}xp(1/\theta)$ de densité égale à :

$$f_{D_i}(d_i;\theta) = \frac{1}{\theta}\exp\left(-\frac{d_i}{\theta}\right), \quad \forall d_i \in \mathbb{R}^+ \quad (10.186)$$

Cette fonction satisfait les hypothèses de régularité. On note $\widehat{\theta} = n^{-1}\sum_{i=1}^{n} D_i$ l'estimateur du maximum de vraisemblance. Déterminons la loi asymptotique de cet estimateur. Soit θ_0 la vraie valeur du paramètre. Nous avons vu que dans ce cas, la quantité d'information de Fisher moyenne et la quantité d'information de Fisher associée à l'échantillon sont respectivement

définies par :

$$\mathrm{I}(\theta_0) = \frac{1}{\theta_0^2} \qquad \mathrm{I}_n(\theta_0) = \frac{n}{\theta_0^2} \tag{10.187}$$

Puisque le problème est régulier, l'estimateur $\widehat{\theta}$ est asymptotiquement normalement distribué :

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, \mathrm{I}^{-1}(\theta_0)\right) \tag{10.188}$$

soit dans notre cas :

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, \theta_0^2\right) \tag{10.189}$$

Ce résultat implique que pour une taille d'échantillon n suffisamment importante, mais finie (► chapitre 9) :

$$\widehat{\theta} \overset{asy}{\approx} \mathcal{N}\left(\theta_0, \frac{\theta_0^2}{n}\right) \tag{10.190}$$

La variance asymptotique de $\widehat{\theta}$ est égale à :

$$\mathbb{V}_{asy}\left(\widehat{\theta}\right) = \frac{\theta^2}{n} = \mathrm{I}_n^{-1}(\theta_0) \tag{10.191}$$

L'estimateur est asymptotiquement efficace.

Les points **clés**

→ Dans le cas discret, la fonction de vraisemblance d'un échantillon correspond à la probabilité jointe d'observation d'un échantillon (réalisation).

→ Dans le cas continu, la fonction de vraisemblance d'un échantillon correspond à la densité jointe des variables de l'échantillon.

→ L'estimateur du maximum de vraisemblance est la quantité qui maximise la log-vraisemblance.

→ Sous les hypothèses de régularité, l'estimateur du maximum de vraisemblance est convergent, efficace et asymptotiquement normalement distribué.



3 questions à

Alina **Catargiu**

Chef de Projet Risque, Crédit Agricole Consumer Finance



Quel est votre parcours professionnel et votre mission actuelle chez Crédit Agricole Consumer Finance ?

À l'issue de mon master d'Économétrie et de Statistique appliquée à l'Université d'Orléans et de mon stage chez Sofinco, j'ai été embauchée en 2010 chez Crédit Agricole Consumer Finance. Je travaille actuellement au sein du Pôle Prévention du Risque et Innovation qui est rattaché à la direction Crédit France. La mission principale de ce pôle est d'assurer la maîtrise du risque au regard des objectifs fixés par la direction générale, en définissant et en mettant en œuvre les politiques d'acceptation des crédits et de gestion de la fraude. Mon métier se concentre autour du développement d'outils de sélection du risque (scores, règles d'acceptation), de la recherche méthodologique et de la création d'outils génériques de suivi des scores. Je participe également au contrôle de la qualité des scores et des études réalisés au sein de mon équipe.

Dans le cadre de votre activité, utilisez-vous la méthode d'estimation du maximum de vraisemblance ?

Chez Crédit Agricole Consumer Finance, l'objectif principal des modèles statistiques est de détecter les groupes d'individus à risque lors de la souscription d'un crédit. Nous travaillons essentiellement avec des modèles de scoring, c'est-à-dire des modèles statistiques qui permettent d'attribuer un score reflétant le risque associé à un client ou à un client potentiel en fonction de ses caractéristiques individuelles. Le modèle de scoring le plus utilisé est le modèle de régression logistique ou modèle logit. Le logit permet d'estimer la probabilité conditionnelle de défaut en fonction des caractéristiques individuelles du client. Chacune de ces caractéristiques est associée à un paramètre à estimer par la méthode du maximum de vraisemblance.

Est-ce que cette méthode d'estimation est programmée dans les logiciels professionnels que vous utilisez chez Crédit Agricole Consumer Finance ?

Tous nos travaux de modélisation statistique sont réalisés sous SAS. L'estimation des paramètres d'une régression logistique se fait *via* la procédure LOGISTIC qui utilise comme méthode d'estimation le maximum de vraisemblance. Dans le cadre d'un modèle logit, il n'y a pas de solution analytique pour définir l'estimateur du maximum de vraisemblance. Le logiciel utilise donc une méthode d'optimisation numérique (méthode du score de Fisher ou méthode de Newton-Raphson). ■

ÉVALUATION

► Corrigés sur www.dunod.com

QCM

Pour chacune des questions suivantes, indiquez si les affirmations sont vraies ou fausses (il peut y avoir plusieurs réponses vraies pour chaque question).

1 Vraisemblance et log-vraisemblance

a. La log-vraisemblance d'un échantillon est égale au logarithme de la vraisemblance de l'échantillon.

b. Pour une variable continue, la vraisemblance d'un échantillon correspond à la probabilité jointe d'apparition d'un échantillon.

c. La log-vraisemblance d'un échantillon est égale au produit des vraisemblances individuelles associées à chaque observation de cet échantillon.

d. La log-vraisemblance d'un échantillon est une variable aléatoire.

e. La vraisemblance d'un échantillon dépend de deux arguments : le vecteur de paramètres et les données de l'échantillon.

2 Estimateur du maximum de vraisemblance

a. L'estimateur du maximum de vraisemblance est une constante.

b. L'estimateur du maximum de vraisemblance est une fonction des variables aléatoires de l'échantillon.

c. L'estimateur du maximum de vraisemblance est la solution du programme de maximisation de la log-vraisemblance.

d. Le gradient de l'échantillon, évalué au point de la réalisation de l'estimateur du maximum de vraisemblance, est égal à zéro.

e. Si l'on souhaite estimer trois paramètres, la hessienne associée à l'échantillon est un vecteur de dimension 3×1.

3 Score, hessienne et information de Fisher

a. Le gradient est une réalisation du score.

b. L'espérance du score est nulle.

c. L'information de Fisher associée à l'échantillon est égale à la variance du score de l'échantillon.

d. L'information de Fisher moyenne est égale à l'information de Fisher de l'échantillon divisée par la taille de celui-ci.

e. L'information de Fisher moyenne correspond à l'information de Fisher associée à une observation de l'échantillon.

4 Propriétés de l'estimateur du maximum de vraisemblance

a. L'estimateur du maximum de vraisemblance est toujours sans biais.

b. Sous les hypothèses de régularité usuelles, l'estimateur du maximum de vraisemblance est convergent au sens fort.

c. Sous les hypothèses de régularité, l'estimateur du maximum de vraisemblance est asymptotiquement normalement distribué.

d. Sous les hypothèses de régularité, l'estimateur du maximum de vraisemblance a une variance asymptotique inférieure (dans le cas scalaire) à la borne FDCR.

e. La variance asymptotique de l'estimateur du maximum de vraisemblance est égale à la matrice d'information de Fisher associée à l'échantillon.

Sujets d'examen

5 Maximum de vraisemblance (HEC Lausanne, 2014)

On considère un n-échantillon $(X_1,\ldots,X_n)$ de variables aléatoires réelles, continues et i.i.d. de même loi que X. On suppose que X a une distribution log-normale de paramètres μ et σ^2 :

$$X \overset{i.i.d.}{\sim} \ln \mathcal{N}\left(\mu,\sigma^2\right) \qquad (10.192)$$

De façon équivalente, la variable $\ln(X)$ admet une distribution normale :

$$\ln(X) \overset{i.i.d.}{\sim} \mathcal{N}\left(\mu,\sigma^2\right) \quad (10.193)$$

La fonction de densité de la variable X est donnée par :

$$f_X\left(x;\mu,\sigma^2\right) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right) \quad \forall x \in \mathbb{R}^+ \quad (10.194)$$

On suppose que le paramètre μ est connu et l'on cherche à estimer le paramètre σ^2. Pour simplifier les calculs, on pourra poser $\theta = \sigma^2$.

1. Déterminer la log-vraisemblance associée à la réalisation de l'échantillon $(x_1,\ldots,x_n)$.

2. Déterminer l'estimateur $\widehat{\sigma}^2$ du maximum de vraisemblance du paramètre σ^2.

3. Montrer que l'estimateur $\widehat{\sigma}^2$ est sans biais.

4. Montrer que l'estimateur $\widehat{\sigma}^2$ est convergent (au sens faible).

5. Déterminer le score associé à l'échantillon. Montrer que son espérance est nulle pour toute valeur de σ^2.

6. Déterminer la quantité d'information de Fisher associée à l'échantillon et la quantité moyenne d'information de Fisher.

7. Déterminer la loi asymptotique de l'estimateur $\widehat{\sigma}^2$.

8. Montrer que l'estimateur $\widehat{\sigma}^2$ est efficace au sens de la borne FDCR.

9. Proposer un estimateur convergent de la variance asymptotique de $\widehat{\sigma}^2$.

6 Maximum de vraisemblance (Université d'Orléans, 2012)

Soit X une variable aléatoire continue positive distribuée selon une loi Gamma de paramètres α et β, notée $\Gamma(\alpha,\beta)$, admettant pour fonction de densité :

$$f_X(x;\alpha,\beta) = \frac{x^{\alpha-1}\exp\left(-\frac{x}{\beta}\right)}{\Gamma(\alpha)\,\beta^{\alpha}} \quad \forall x \in [0,+\infty[\quad (10.195)$$

où $\Gamma(\alpha)$ désigne la fonction Gamma, avec

$$\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1}\exp(-t)\,dt \quad (10.196)$$

On admet que

$$\mathbb{E}(X) = \alpha\beta, \quad \mathbb{V}(X) = \alpha\beta^2 \quad (10.197)$$

On suppose que le paramètre α est connu et que le paramètre β est inconnu avec $\alpha > 0$ et $\beta > 0$. On souhaite estimer le paramètre β par la méthode du maximum de vraisemblance à partir d'un n-échantillon $(X_1,\ldots,X_n)$ i.i.d. de même loi que X.

1. Déterminer la log-vraisemblance associée à la réalisation de l'échantillon $(x_1,\ldots,x_n)$.

2. Montrer que le gradient $g_n(\beta;x)$ et la hessienne $H_n(\beta;x)$ associés à cette vraisemblance s'écrivent :

$$g_n(\beta;x) = \frac{1}{\beta^2}\sum_{i=1}^{n} x_i - \frac{n\alpha}{\beta} \quad (10.198)$$

$$H_n(\beta;x) = -\frac{2}{\beta^3}\sum_{i=1}^{n} x_i + \frac{n\alpha}{\beta^2} \quad (10.199)$$

3. Déterminer le score associé à l'échantillon. Montrer que son espérance est nulle pour toute valeur de β.

4. Déterminer l'estimateur $\widehat{\beta}$ du maximum de vraisemblance du paramètre β.

5. Montrer que l'estimateur du maximum de vraisemblance $\widehat{\beta}$ est sans biais.

6. Montrer que l'estimateur $\widehat{\beta}$ est convergent.

7. L'estimateur du maximum de vraisemblance $\widehat{\beta}$ est-il efficace au sens de la borne FDCR ?

8. En utilisant le théorème central limite déterminer la loi asymptotique de la quantité $n^{-1}\sum_{i=1}^{n} X_i$. En déduire la loi asymptotique de l'estimateur $\widehat{\beta}$.

9. Retrouver la loi asymptotique de l'estimateur $\widehat{\beta}$ de la question 8 en utilisant les propriétés asymptotiques du maximum de vraisemblance.

10. On considère un échantillon de taille $n = 10$ de variables $X_1,\ldots,X_n$ distribuées selon une loi $\Gamma(2,\beta)$, pour lequel on observe les réalisations suivantes :

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
3,5	5,2	2,1	6,3	4,7	3,5	3,0	4,9	2,1	4,7

Proposer une estimation ponctuelle du paramètre β.

Chapitre 11

Selon une légende du marketing, la chaîne de distribution américaine Wall-Mart aurait mis en place dans les années 1980 une réorganisation de ses rayons visant à présenter côte à côte les couches pour bébés et les packs de bières. Cette réorganisation aurait fait suite à un constat simple : l'examen de millions de tickets de caisse montrerait que les ventes des deux produits sont statistiquement liées. Au-delà de la légende, comment parvenir à une telle conclusion ? On formalise généralement ce type de problème sous la forme d'un test statistique : on teste une hypothèse dite nulle (les ventes ne sont pas liées par exemple) et l'on construit une région critique. Une région critique est une règle de décision concernant le rejet ou le non-rejet de l'hypothèse nulle. L'avantage d'une telle démarche est qu'elle permet de contrôler les risques associés à la décision. C'est pourquoi les tests statistiques sont aujourd'hui si souvent employés dans de très nombreux domaines d'activité économique.

LES GRANDS AUTEURS



Jerzy Neyman (1894-1981)



Jerzy Neyman est considéré comme l'un des grands fondateurs de la théorie statistique moderne, avec notamment Karl Pearson (1857-1936) dont il fut l'étudiant. Ses travaux ont largement contribué à la théorie moderne des probabilités et à la théorie des tests.

Il étudie en Pologne, puis à Londres où il travaille avec Egon Sharpe Pearson, le fils de Karl Pearson. Ensemble, ils développent en 1928 le test dit du rapport de vraisemblance.

Mais c'est en 1933 qu'ils apportent la démonstration que ce test est le plus puissant grâce au fameux **lemme de Neyman-Pearson** que nous étudierons dans ce chapitre. Jerzy Neyman émigre aux Etats-Unis en 1937. Il fonde alors le département de statistiques de la prestigieuse université de Berkeley en Californie. ■

Théorie des tests

Plan

Pré-requis

→ **Connaître** les différentes notions de convergence (▸ chapitre 8).

→ **Connaître** la notion d'estimateur (▸ chapitre 9).

Objectifs

→ **Comprendre** les notions d'hypothèses nulle et alternative.

→ **Comprendre** les notions de statistique de test et de valeur critique.

→ **Comprendre** les notions de niveau et de puissance d'un test.

→ **Construire** la région critique d'un test.

→ **Apprendre à conclure** quant au rejet de l'hypothèse nulle.

La théorie des tests (ou inférence) étudie la construction et les propriétés des tests statistiques. Un test statistique est une règle de décision permettant de rejeter ou de ne pas rejeter une hypothèse, appelée hypothèse nulle, en fonction des observations d'un échantillon. La théorie des tests est la théorie fondamentale de ce que l'on appelle aujourd'hui la statistique décisionnelle ou *business intelligence*, et de ce que l'on appelait autrefois la statistique mathématique. Elle est le fondement de tous les outils statistiques modernes d'aide à la décision. Au-delà de la règle de décision, le principal avantage d'un test statistique est qu'il permet de mesurer ou de contrôler les risques associés à cette décision. C'est pourquoi ces tests sont très utilisés en pratique.

De nos jours, les applications des tests statistiques en économie et en gestion sont omniprésentes dans notre vie quotidienne et dans la vie des entreprises. Par exemple, lorsque vous sollicitez un prêt à la consommation ou un prêt immobilier auprès d'une banque, celle-ci cherche à déterminer si vous êtes un bon client, c'est-à-dire si vous serez apte à rembourser le capital et les intérêts dans le futur. Bien évidemment, rien n'indique *a priori* que vous soyez un client à risque ou non.

Une façon formelle de répondre à cette question consiste à tester l'**hypothèse nulle** selon laquelle vous êtes un bon client, et à définir une règle de décision du type rejet (ce qui implique un rejet du prêt) ou non-rejet (ce qui implique une acceptation du prêt) de cette hypothèse nulle. C'est le principe général des méthodes de scoring appliquées dans le domaine bancaire, de l'assurance, du marketing, de la détection des fraudes sur internet, etc. Un autre exemple est celui d'une entreprise qui souhaite analyser l'impact d'une campagne marketing à partir d'un échantillon des ventes. Elle formalise ce problème sous la forme d'une hypothèse nulle (la campagne n'a pas eu d'impact), et construit une règle de décision à partir de l'échantillon permettant de conclure quant au rejet ou non de cette hypothèse. Une autre application, très utilisée en marketing, est celle des **tests d'indépendance**, grâce auxquels il est possible de tester si les ventes de deux produits sont liées ou si au contraire elles sont indépendantes (hypothèse nulle).

1 Définitions

Définition 11.1

> Un **test statistique** (ou test) est une **règle de décision** relative à une hypothèse sur la distribution d'une variable d'intérêt dans la population, qui se fonde sur les observations d'un échantillon.

Dans la vie quotidienne, il est généralement relativement facile de se fixer soi-même, ou de se voir conseiller, une règle de décision. Par exemple, si l'hypothèse considérée est celle de la pertinence d'un investissement dans un titre financier, votre conseiller financier peut vous proposer une règle de décision du type : « Achetez si le prix de ce titre descend en dessous de 10 euros ». Le problème avec ce type de règle heuristique, c'est que l'on ne contrôle pas les **risques** associés à la décision. Le fameux adage les « conseilleurs ne sont pas les payeurs » rend bien compte de cette déconnection

qui apparaît parfois entre la règle de décision et le risque encouru. C'est la principale différence avec une règle de décision statistique : un test statistique est une procédure qui permet de **contrôler** ou de **minimiser**, suivant les cas, les risques associés à la décision. C'est pourquoi, un test statistique est toujours associé à trois éléments :

1. Une **hypothèse** nulle et une hypothèse alternative.
2. Une **région critique** fondée sur une statistique de test et une valeur critique.
3. Des **risques** de première espèce et de seconde espèce.

Nous allons désormais présenter ces différents éléments.

1.1 Hypothèses nulle et alternative

Un test, en tant que règle de décision, se réfère toujours à une hypothèse de référence, dite hypothèse nulle.

Définition 11.2

Une hypothèse est une assertion concernant la population. Un test statistique permet de tester la validité d'une hypothèse de référence (ou de base), dite **hypothèse nulle**, contre une **hypothèse alternative**. Ces hypothèses sont respectivement notées H_0 et H_1.

Considérons l'exemple d'un test concernant l'effet d'un traitement médical. Dans ce cas, on peut construire un test de l'hypothèse nulle H_0 : « le traitement n'a pas d'effet » contre une hypothèse alternative H_1 : « le traitement a un effet ». Mais on peut aussi tester l'inverse, c'est-à-dire H_0 : « le traitement a un effet » contre H_1 : « le traitement n'a pas d'effet ». De façon générale, comment choisir l'hypothèse nulle ?

Dans la vie économique, on utilise de nombreux tests statistiques. Pour ces tests, il est généralement possible de calculer les coûts associés au renoncement de l'une ou l'autre des deux hypothèses. La règle est que l'*hypothèse nulle* est celle dont *le coût d'une erreur associée est le plus important*. Par exemple, dans le cadre d'un problème de scoring on cherche à tester si le client présente un risque (risque de défaut, risque de désabonnement, risque de ne pas acheter le produit, etc.) ou si, au contraire, le client n'est pas risqué. Contrairement à une intuition première, le coût associé à une erreur de décision sur l'hypothèse « le client est risqué » est généralement plus faible que le coût associé à une erreur sur l'hypothèse « le client n'est pas risqué ». Prenons l'exemple d'un scoring bancaire relatif à l'attribution d'un prêt immobilier de 200 000 euros sur 20 ans. Si l'on juge que le client est risqué alors qu'il ne l'est pas, la banque ne lui attribue pas le prêt. La banque perd alors l'intégralité des intérêts, soit suivant le taux d'intérêt, environ 100 000 à 150 000 euros. Si au contraire le client est jugé non risqué alors qu'il l'est réellement, la banque lui attribue le prêt et observera sans doute un défaut (partiel ou total) de remboursement dans les années à venir. Mais si ce défaut intervient dans 20 ans, la banque ne perd rien, le capital ayant été remboursé et les intérêts payés. Si ce défaut intervient plus tôt, la banque enregistre des pertes.

Même dans le cas le plus défavorable d'un défaut total intervenant le lendemain de la signature du prêt, ce coût est largement inférieur à celui du premier type de risque.

En effet, dans ce cas la banque saisit le bien immobilier et le revend immédiatement avec une décote et des frais annexes. Ainsi, ses pertes peuvent être limitées à quelques dizaines de milliers d'euros. Par conséquent, le coût associé à une erreur de décision sur l'hypothèse « le client n'est pas risqué » est le plus élevé : c'est cette hypothèse qui sera choisie comme hypothèse nulle.

On distingue deux grandes familles de tests statistiques :

1. Les **tests paramétriques** : ces tests portent sur la valeur d'un ou de plusieurs paramètres de la distribution dans la population d'une variable d'intérêt.
2. Les **tests non-paramétriques** : ces tests portent sur la distribution, les moments (espérance, variance, etc.) ou certaines caractéristiques (l'indépendance par exemple) d'une ou de plusieurs variables aléatoires.

Dans ce chapitre nous étudierons tout d'abord les tests paramétriques (▶ section 3) avant de présenter quelques exemples de tests non-paramétriques (▶ section 4).

Dans le cas des tests paramétriques, l'hypothèse nulle et l'hypothèse alternative portent sur la **valeur d'un paramètre** θ (ou de plusieurs paramètres) de la distribution d'une variable aléatoire (discrète ou continue) X définie sur un univers probabilisé $(X(\Omega), \mathcal{F}, \Pr)$ et admettant une fonction de densité ou une fonction de masse $f_X(x;\theta)$.

Exemple

On admet qu'une variable aléatoire discrète X, définie sur $X(\Omega) = \mathbb{N}$, suit une loi de Poisson de paramètre θ, où θ est un paramètre réel positif inconnu, telle que :

$$f_X(x;\theta) = \Pr(X = x) = \exp(-\theta)\frac{\theta^x}{x!} \quad \forall x \in \mathbb{N} \tag{11.1}$$

On cherche à tester l'hypothèse nulle $H_0 : \theta = 2$ contre une hypothèse alternative $H_1 : \theta = 3$.

Parmi les tests paramétriques, on distingue les tests d'hypothèses simples et les tests d'hypothèses composites.

Définition 11.3

Une **hypothèse simple** caractérise complétement la distribution de la variable d'intérêt. Une **hypothèse composite** ne permet pas de caractériser la distribution de la variable d'intérêt.

Exemple

Considérons une variable aléatoire X distribuée selon une loi de Student $t(\theta)$ où $\theta > 0$ est un paramètre inconnu. L'hypothèse nulle $H_0 : \theta = 2$ est une hypothèse simple, car sous H_0 on connaît exactement la loi de la variable X, *i.e.* $X \sim t(2)$. Les hypothèses nulles $H_0 : \theta > 2$, $H_0 : \theta < 2$ ou $H_0 : \theta \neq 2$ sont des hypothèses composites. En effet, on ne sait pas quelle est la loi exacte de X sous H_0. Pour $H_0 : \theta > 2$, cette distribution peut être, par exemple, $X \sim t(3)$, $X \sim t(10)$ ou $X \sim t(100)$.

Remarque : On peut construire des tests d'une hypothèse simple contre une hypothèse simple. Par exemple, $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$. On peut construire différents tests d'une hypothèse simple contre une hypothèse composite. Par exemple, $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$. On peut aussi construire, même si c'est plus rare, des tests d'une hypothèse composite contre une hypothèse simple ou une hypothèse composite, du type $H_0 : \theta < \theta_0$ contre $H_1 : \theta = \theta_1$ ou $H_0 : \theta < \theta_0$ contre $H_1 : \theta > \theta_0$.

Parmi les tests d'une hypothèse simple contre une hypothèse composite, on distingue les **tests unilatéraux** des **tests bilatéraux**. Cette distinction sera particulièrement importante pour la définition de la région critique. Le terme unilatéral renvoie au fait que sous H_1, la valeur du paramètre θ ne peut être que supérieure (ou inférieure suivant le test) à la valeur de θ sous l'hypothèse nulle H_0 : la valeur de θ ne prend qu'une seule « direction ». Le terme bilatéral signifie, qu'au contraire, la valeur de θ sous l'hypothèse alternative est différente (inférieure ou supérieure) de la valeur sous l'hypothèse nulle.

Définition 11.4

Un **test unilatéral gauche** est un test de la forme $H_0 : \theta = \theta_0$ contre $H_1 : \theta < \theta_0$.
Un **test unilatéral droit** est un test de la forme $H_0 : \theta = \theta_0$ contre $H_1 : \theta > \theta_0$.

Définition 11.5

Un **test bilatéral** est un test de la forme $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$.

Enfin, signalons que lorsqu'un test porte sur plusieurs paramètres, on parle de test d'hypothèses jointes ou de **test joint**.

Définition 11.6

Un **test d'hypothèses jointes** (ou test joint) est un test dont l'hypothèse nulle porte sur plusieurs paramètres $\theta_1,\ldots,\theta_k$ de la distribution de la variable d'intérêt.

$$H_0 : \theta_1 = a_1 \text{ et } \theta_2 = a_2 \text{ et } \ldots \text{ et } \theta_k = a_k \tag{11.2}$$

Par exemple, si la variable d'intérêt X vérifie $X \sim \mathcal{N}\left(m,\sigma^2\right)$, on peut construire un test joint de la forme $H_0 : m = m_0$ et $\sigma^2 = \sigma_0^2$. L'hypothèse alternative peut s'écrire sous la forme $H_1 : m \neq m_0 \; et \; \sigma^2 \neq \sigma_0^2$ ou sous la forme $H_1 : m \neq m_0 \; ou \; \sigma^2 \neq \sigma_0^2$.

1.2 Région critique

Supposons que l'on dispose d'un n-échantillon $(X_1,\ldots,X_n)$ de variables i.i.d. de même loi que la variable X. Comment tester, à partir de cet échantillon, l'hypothèse nulle d'un test paramétrique portant sur la valeur d'un paramètre θ de sa distribution dans la population ? Pour ce faire, nous allons construire une **région critique** à partir de deux éléments : une **statistique de test**[1] et une **valeur critique**. Commençons par définir la notion de statistique de test.

Définition 11.7

Une **statistique de test**, notée T_n, est une **variable aléatoire** définie comme une fonction des variables de l'échantillon $X_1,\ldots,X_n$:

$$T_n\left(X_1,\ldots,X_n\right) \tag{11.3}$$

1 Nous avons déjà défini la notion de *statistique descriptive* dans le chapitre 1 et dans le chapitre 9, consacré à l'estimation. La définition d'une *statistique de test* est similaire.

Remarque : En général, mais pas toujours, la statistique de test correspond à un estimateur $\widehat{\theta}$ du paramètre θ ou à une variable transformée de cet estimateur.

Une statistique de test étant une variable aléatoire, on peut en caractériser la **distribution** (ou distribution d'échantillonnage). Comme pour un estimateur, on distingue la loi exacte d'une statistique de test, valable pour toute valeur de n (cette loi exacte est généralement difficile à dériver sauf dans des cas simples), de la loi asymptotique, valable pour une taille d'échantillon n suffisamment grande, mais finie.

$$T_n(X_1,\ldots,X_n) \sim \text{loi exacte } \forall n \in \mathbb{N} \tag{11.4}$$

$$T_n(X_1,\ldots,X_n) \overset{asy}{\approx} \text{loi asymptotique} \tag{11.5}$$

La région critique correspond à la règle de décision du test statistique. Cette règle est extrêmement simple : si la réalisation de la statistique de test, obtenue à partir des observations $(x_1,\ldots,x_n)$ appartient à la région critique, on rejette l'hypothèse nulle H_0. La région critique est un ensemble délimité par une ou des **valeurs critiques**, suivant les cas.

Définition 11.8

La **région critique** d'un test, notée W, est un ensemble de réalisations de la statistique de test (ou de façon équivalente un ensemble d'échantillons) pour lesquelles l'hypothèse nulle du test est **rejetée** :

$$W = \{x_1,\ldots,x_n : T_n(x_1,\ldots,x_n) \in \Gamma(c)\} \tag{11.6}$$

où $(x_1,\ldots,x_n)$ désigne un échantillon, $T_n(x_1,\ldots,x_n)$ la réalisation associée de la statistique de test, et $\Gamma(c)$ un ensemble délimité par une (ou plusieurs) valeur(s) critique(s), notée(s) c.

Exemple

Voici quelques exemples de formes de régions critiques usuelles :

$$W = \{x_1,\ldots,x_n : T_n(x_1,\ldots,x_n) > c\} \tag{11.7}$$

$$W = \{x_1,\ldots,x_n : c_1 < T_n(x_1,\ldots,x_n) < c_2\} \tag{11.8}$$

$$W = \{x_1,\ldots,x_n : |T_n(x_1,\ldots,x_n)| > c\} \tag{11.9}$$

où c, c_1 et c_2 sont des valeurs critiques, généralement déterminées à partir de tables statistiques (▶ chapitre 7).

Remarque : Par souci de simplicité dans les notations, nous noterons la région critique sous la forme W= $\{x : T_n(x) \in \Gamma(c)\}$, où x en minuscule renvoie à la réalisation du n-échantillon $(x_1,\ldots,x_n)$.

Ainsi, la procédure d'un test est la suivante : on calcule la réalisation de la statistique de test à partir de l'échantillon d'observations. Si cette réalisation appartient à la région critique, on rejette l'hypothèse nulle H_0. Par conséquent, un test statistique est une règle de décision qui spécifie :

– l'ensemble des échantillons pour lesquels on **rejette** H_0 ;

– l'ensemble des échantillons pour lesquels **on ne peut pas rejeter** H_0.

Remarque : La seule conclusion que l'on peut tirer d'un test, c'est celle du *rejet* ou du *non-rejet* de l'hypothèse nulle H_0. On ne doit jamais tirer d'un test des conclusions du type « on accepte H_0 », « on accepte H_1 », « on rejette H_1 », etc.

Définition 11.9

La région complémentaire de la région critique est appelée **région de non-rejet** de l'hypothèse nulle H_0, notée $\overline{W}$, telle que :

$$\overline{W} = \{x : T_n(x) \notin \Gamma(c)\} \quad (11.10)$$

Si la réalisation de la statistique de test appartient à la zone de non-rejet, on conclut au non-rejet de l'hypothèse nulle H_0.

1.3 Risques

Lorsque l'on considère un test statistique, on ne peut prendre que l'une ou l'autre des deux décisions suivantes : soit on rejette l'hypothèse nulle H_0, soit on ne rejette pas l'hypothèse nulle H_0. L'avantage principal d'un test statistique par rapport à une règle de décision heuristique (une décision au hasard par exemple), est qu'il permet de contrôler les risques associés à la décision.

Quels sont ces risques ? Comme l'indique le tableau 11.1, on distingue deux types de risque : le **risque de première espèce** et le **risque de deuxième espèce**. Dans ce tableau, on croise la décision (rejet ou non-rejet de H_0) et la validité de H_0 ou de H_1 dans la population. Ainsi, si l'on rejette H_0 alors que H_1 est vraie ou si l'on ne rejette pas H_0 alors que H_0 est vraie, on ne commet pas d'erreur. En revanche, si l'on rejette H_0 alors que H_0 est vraie, on commet une erreur dite de type I ou de première espèce. Si on ne rejette pas H_0 alors que H_1 est vraie, on commet une erreur dite de type II ou de deuxième espèce.

▼ Tableau 11.1 **Risque I et risque II**

		Décision	
		Non-rejet de H_0	**Rejet de H_0**
Population	**H_0 vraie**	Décision correcte	Erreur de type I
	H_1 vraie	Erreur de type II	Décision correcte

Définition 11.10

Le risque I ou **risque de première espèce**, correspond au risque de rejeter l'hypothèse nulle H_0 alors qu'elle est effectivement vraie dans la population.

Définition 11.11

Le risque II ou **risque de seconde espèce** correspond au risque de ne pas rejeter l'hypothèse nulle H_0 alors que l'hypothèse alternative H_1 est valide dans la population.

Pour une règle de décision donnée, c'est-à-dire pour une région critique W, on cherche à quantifier les probabilités associées à ces deux types de risque.

Définition 11.12

Le **niveau** (ou la taille) d'un test correspond à la probabilité associée au risque de première espèce. Par convention, cette probabilité est notée α :

$$\alpha = \Pr(W | H_0) \tag{11.11}$$

où W désigne la région critique du test.

Ainsi, le niveau correspond à la probabilité de rejeter H_0, c'est-à-dire d'être dans la région critique W, sachant que l'hypothèse nulle H_0 est vraie dans la population. C'est donc la probabilité de rejeter à tort l'hypothèse nulle H_0. Bien évidemment, plus le niveau d'un test est faible, plus la probabilité d'erreur de première espèce est faible et mieux c'est. Le symbole $|H_0$ signifie « sachant que H_0 est vraie ».

Remarque : Pour un test d'hypothèse nulle composite, le niveau du test devient : $\alpha = \sup_{\theta_0 \in H_0} \Pr(W | H_0)$.

De façon similaire, nous pouvons définir la probabilité associée au risque de deuxième espèce et la puissance d'un test, définie comme le complémentaire de cette probabilité.

Définition 11.13

La **puissance** d'un test correspond à la probabilité de rejet de l'hypothèse nulle H_0 alors que l'hypothèse alternative H_1 est vraie :

$$\text{Puissance} = \Pr(W | H_1) = 1 - \beta \tag{11.12}$$

où β correspond à la probabilité de l'erreur de deuxième espèce, *i.e.* $\beta = \Pr\left(\overline{W} \middle| H_1\right)$ et où $\overline{W}$ désigne la région de non-rejet.

La puissance correspond à la probabilité d'être dans la région critique (et donc de rejeter l'hypothèse nulle) alors que l'hypothèse alternative H_1 est vraie dans la population. Par conséquent, plus un test est puissant, plus la probabilité d'erreur de deuxième espèce est faible et mieux c'est.

Propriété

Détermination du niveau et de la puissance

Afin de caractériser le *niveau* d'un test, on doit utiliser la distribution de la statistique de test $T_n(X)$ sous l'hypothèse nulle H_0. Il peut s'agir soit de la loi exacte, soit de la loi asymptotique :

$$T_n(X) \underset{H_0}{\sim} D \quad \text{ou} \quad T_n(X) \underset{H_0}{\overset{asy}{\approx}} D \tag{11.13}$$

Afin de caractériser la probabilité de risque II, β, ou la *puissance*, on doit utiliser la distribution de la statistique de test $T_n(X)$ sous l'hypothèse alternative H_1 :

$$T_n(X) \underset{H_1}{\sim} D \quad \text{ou} \quad T_n(X) \underset{H_1}{\overset{asy}{\approx}} D \tag{11.14}$$

Appliquons ces définitions dans le cadre d'un exemple.

Exemple

On considère un n-échantillon $(X_1,\ldots,X_n)$, avec $n = 100$, de variables aléatoires i.i.d. telles que $X_i \sim \mathcal{N}\left(m,\sigma^2\right)$ où m est un paramètre inconnu et $\sigma^2 = 1$. On souhaite tester :

$$\mathrm{H}_0 : m = m_0 = 1{,}2 \quad \text{contre} \quad \mathrm{H}_1 : m = m_1 = 1 \tag{11.15}$$

Un économètre vous propose une région critique de la forme :

$$\mathrm{W} = \{x : \overline{x}_n < c\} \tag{11.16}$$

où $\overline{x}_n$ désigne la réalisation de la moyenne empirique $\overline{X}_n = n^{-1}\sum_{i=1}^{n} X_i$ et c est une constante (valeur critique) égale à 1,0718. Cette région critique s'interprète de la façon suivante : si la réalisation de la moyenne empirique est inférieure à 1,0718, on rejette l'hypothèse nulle H_0 : $m = 1{,}2$. Calculons la *taille* et la *puissance* de ce test. Sous l'hypothèse nulle $\mathrm{H}_0 : m = m_0$, la loi exacte de la moyenne empirique $\overline{X}_n$ (statistique de test) est (▶ chapitre 9) :

$$\overline{X}_n \underset{\mathrm{H}_0}{\sim} \mathcal{N}\left(m_0, \frac{\sigma^2}{n}\right) \Longleftrightarrow \frac{\overline{X}_n - m_0}{\sigma/\sqrt{n}} \underset{\mathrm{H}_0}{\sim} \mathcal{N}(0,1) \tag{11.17}$$

Par conséquent, la taille du test est égale à :

$$\alpha = \Pr(\mathrm{W}|\,\mathrm{H}_0) = \Pr\left(\overline{X}_n < c\middle|\,\mathrm{H}_0\right) \tag{11.18}$$

$$= \Pr\left(\frac{\overline{X}_n - m_0}{\sigma/\sqrt{n}} < \frac{c - m_0}{\sigma/\sqrt{n}}\middle|\,\mathrm{H}_0\right) \tag{11.19}$$

$$= \Phi\left(\frac{c - m_0}{\sigma/\sqrt{n}}\right) \tag{11.20}$$

où $\Phi(.)$ désigne la fonction de répartition de la loi normale centrée réduite. D'après les données de l'énoncé, la taille du test est égale à :

$$\alpha = \Phi\left(\frac{1{,}0718 - 1{,}2}{1/\sqrt{100}}\right) = \Phi(-1{,}2816) = 0{,}10 \tag{11.21}$$

Ainsi, avec la règle de décision associée à la région critique W= $\{x : \overline{x}_n < 1{,}0718\}$, il y a 10 % de chances de rejeter à tort l'hypothèse nulle $\mathrm{H}_0 : m = 1{,}2$ alors qu'elle est vraie. Sous l'hypothèse alternative, $\mathrm{H}_1 : m = m_1$, la loi exacte de la moyenne empirique $\overline{X}_n$ (statistique de test) est :

$$\overline{X}_n \underset{\mathrm{H}_1}{\sim} \mathcal{N}\left(m_1, \frac{\sigma^2}{n}\right) \Longleftrightarrow \frac{\overline{X}_n - m_1}{\sigma/\sqrt{n}} \underset{\mathrm{H}_1}{\sim} \mathcal{N}(0,1) \tag{11.22}$$

Par conséquent, la puissance du test est égale à :

$$\text{Puissance} = \Pr(\mathrm{W}|\,\mathrm{H}_1) = \Pr\left(\overline{X}_n < c\middle|\,\mathrm{H}_1\right) \tag{11.23}$$

$$= \Pr\left(\frac{\overline{X}_n - m_1}{\sigma/\sqrt{n}} < \frac{c - m_1}{\sigma/\sqrt{n}}\middle|\,\mathrm{H}_1\right) \tag{11.24}$$

$$= \Phi\left(\frac{c - m_1}{\sigma/\sqrt{n}}\right) \tag{11.25}$$

La probabilité de risque de deuxième espèce est égale à :

$$\beta = 1 - \text{Puissance} = 1 - \Phi\left(\frac{c - m_1}{\sigma/\sqrt{n}}\right) \tag{11.26}$$

D'où :

$$\text{Puissance} = \Phi\left(\frac{1{,}0718 - 1}{1/\sqrt{100}}\right) = \Phi(0{,}7184) = 0{,}7638 \tag{11.27}$$

$$\beta = 1 - \text{Puissance} = 1 - 0{,}7638 = 0{,}2362 \tag{11.28}$$

Par conséquent, avec la région critique W= $\{x : \overline{x}_n < 1{,}0718\}$, il y a 23,62 % de chances de ne pas rejeter l'hypothèse nulle $H_0 : m = 1{,}2$ alors que l'hypothèse alternative $H_0 : m = 1$ est vraie.

2 Règle de décision et puissance d'un test

L'objectif de cette section est de présenter la règle de décision d'un test pour un **niveau** de risque de première espèce donné. Nous verrons que cette décision peut aussi être prise sur la base de la valeur p ou **p-value**. Enfin, nous caractériserons la **fonction puissance** d'un test.

2.1 Règle de décision

À partir de ces différents éléments (risque de première espèce et de deuxième espèce, région critique, valeur critique et statistique de test), nous pouvons à présent envisager la mise en œuvre d'un test statistique. Mais pour cela, nous devons lever un problème d'indétermination. Reprenons les résultats de l'exemple précédent portant sur le test de l'espérance d'un échantillon de variables normales. Nous avions obtenu des probabilités de risque I et de risque II, respectivement égales à :

$$\alpha = \Phi\left(\frac{c - m_0}{\sigma/\sqrt{n}}\right), \quad \beta = 1 - \Phi\left(\frac{c - m_1}{\sigma/\sqrt{n}}\right) \tag{11.29}$$

Les valeurs m_0 et m_1 sont fixées par l'utilisateur (hypothèses du test), ainsi que la taille d'échantillon n. On obtient un système à deux équations et trois inconnues : α, β (ou la puissance) et la valeur critique c. Le système est donc indéterminé.

Propriété

Arbitrage risque I/risque II

De façon générale, il existe un **arbitrage** entre le risque de première espèce et le risque de deuxième espèce.

Illustrons cet arbitrage entre le risque de première espèce et le risque de deuxième espèce en fonction de la valeur critique c, par un exemple numérique.

Exemple

On considère un n-échantillon $(X_1,\ldots,X_n)$ de variables aléatoires i.i.d. telles que $X_i \sim \mathcal{N}\left(m,\sigma^2\right)$ où le paramètre m est inconnu. On suppose que $n = 100$ et $\sigma^2 = 1$, et l'on souhaite tester :

$$H_0 : m = m_0 = 1{,}2 \quad \text{contre} \quad H_1 : m = m_1 = 1 \tag{11.30}$$

On admet que la région critique du test est de la forme :

$$W = \{x : \overline{x}_n < c\} \tag{11.31}$$

où la statistique de test $\overline{X}_n$ (moyenne empirique) vérifie :

$$\overline{X}_n \underset{H_0}{\sim} \mathcal{N}\left(m_0, \frac{\sigma^2}{n}\right), \quad \overline{X}_n \underset{H_1}{\sim} \mathcal{N}\left(m_1, \frac{\sigma^2}{n}\right) \tag{11.32}$$

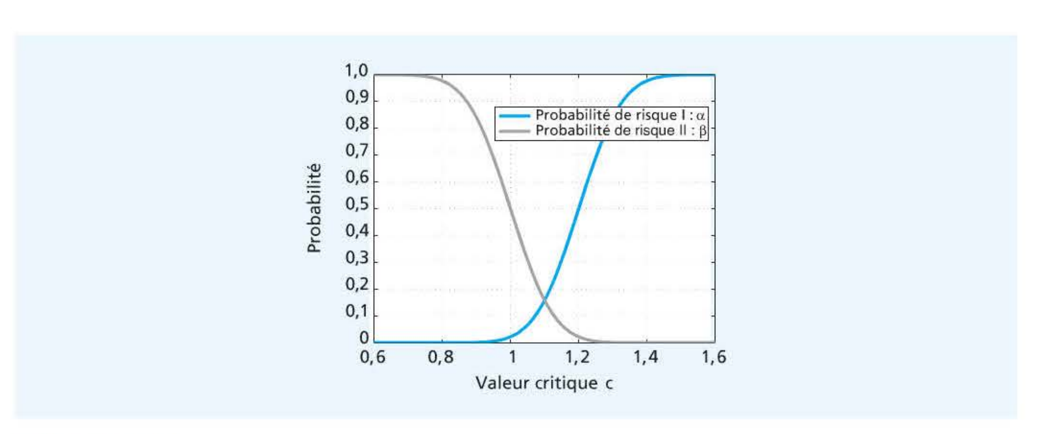
Sous ces hypothèses, les probabilités de risque I et de risque II peuvent s'exprimer en fonction de la valeur critique c comme suit :

$$\alpha = \Pr(W|H_0) = \Phi\left(\frac{c - m_0}{\sigma/\sqrt{n}}\right) = \Phi(10 \times (c - 1{,}2)) \tag{11.33}$$

$$\beta = \Pr\left(\overline{W}\middle|H_1\right) = 1 - \Phi\left(\frac{c - m_1}{\sigma/\sqrt{n}}\right) = 1 - \Phi(10 \times (c - 1)) \tag{11.34}$$

où $\Phi(.)$ désigne la fonction de répartition de la loi normale centrée réduite.

Puisque les fonctions de répartition sont toujours des fonctions strictement croissantes, le niveau α est une fonction croissante de la valeur critique c, tandis que la probabilité de risque de deuxième espèce β est une fonction décroissante de c. Sur la figure 11.1 sont représentées les probabilités α et β en fonction de la valeur critique c. On vérifie que lorsque β augmente, α diminue et *vice et versa*. Cela confirme l'arbitrage entre le risque de première espèce et le risque de deuxième espèce.



▲ **Figure 11.1 Arbitrage entre risque I et risque II**

Par convention, la solution pour résoudre cet arbitrage consiste à fixer la probabilité du risque de première espèce α.

Propriété

Niveau d'un test

Dans la pratique, le niveau α (ou la taille ou le seuil de significativité) du test est **fixé par l'utilisateur**. On en déduit la valeur critique du test ainsi que sa puissance ou de façon équivalente, la probabilité du risque de deuxième espèce.

Pourquoi fixer le niveau α et non pas la probabilité β de risque II (ou la puissance) ? C'est ici que la question du choix de l'hypothèse nulle prend toute son importance (► section 1.1). En effet, nous avons vu que l'hypothèse nulle est celle pour laquelle le coût d'une erreur associée est le plus élevé. Ainsi, contrôler la probabilité α permet à l'utilisateur de contrôler le risque le plus important. C'est pourquoi, dans la pratique, on fixe le niveau d'un test à un seuil relativement faible, typiquement $\alpha = 5\ \%$ ou $\alpha = 10\ \%$.

Exemple

On considère un n-échantillon de variables i.i.d. $(X_1,\ldots,X_n)$, avec $n = 100$, telles que $X_i \sim \mathcal{N}\left(m,\sigma^2\right)$ où m est un paramètre inconnu et $\sigma^2 = 1$. On souhaite tester l'hypothèse suivante :

$$\mathrm{H}_0 : m = m_0 = 1{,}2 \quad \text{contre} \quad \mathrm{H}_1 : m = m_1 = 1 \tag{11.35}$$

Un économètre propose une région critique de la forme :

$$\mathrm{W} = \{x : \overline{x}_n < c\} \tag{11.36}$$

où $\overline{x}_n$ est une réalisation de la moyenne empirique $\overline{X}_n = n^{-1} \sum_{i=1}^{n} X_i$ (statistique de test) et c est une valeur critique. Déterminons cette valeur critique pour un test de niveau $\alpha = 5\ \%$ ainsi que la puissance associée. D'après les résultats de l'exercice précédent, nous savons que :

$$\alpha = \Pr(\mathrm{W}|\,\mathrm{H}_0) = \Phi\left(\frac{c - m_0}{\sigma/\sqrt{n}}\right) \tag{11.37}$$

Appliquons la fonction de répartition inverse $\Phi^{-1}(.)$ aux deux membres de cette égalité afin de déterminer la valeur critique c.

$$\Phi^{-1}(\alpha) = \frac{c - m_0}{\sigma/\sqrt{n}} \Longleftrightarrow c = m_0 + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(\alpha) \tag{11.38}$$

Ainsi, nous obtenons :

$$c = 1{,}2 + \frac{1}{\sqrt{100}} \times \Phi^{-1}(0{,}05) = 1{,}2 + \frac{1}{\sqrt{100}} \times (-1{,}6449) = 1{,}0355 \tag{11.39}$$

La région critique du test de niveau $\alpha = 5\ \%$ est définie par :

$$\mathrm{W} = \{x : \overline{x}_n < 1{,}0355\} \tag{11.40}$$

La puissance du test est égale à :

$$\text{Puissance} = \Pr(\mathrm{W}|\,\mathrm{H}_1) = \Phi\left(\frac{c - m_1}{\sigma/\sqrt{n}}\right) \tag{11.41}$$

Ainsi, nous obtenons :

$$\text{Puissance} = \Phi\left(\frac{1{,}2 - 1}{1/\sqrt{100}} + \Phi^{-1}(0{,}05)\right) = 0{,}6388 \tag{11.42}$$

Avec la région critique W= $\{x : \overline{x}_n < 1{,}0355\}$, il y a 63,88 % de chances de rejeter l'hypothèse nulle $\mathrm{H}_0 : m = 1{,}2$ lorsque l'espérance des variables X_i est égale à $m = 1$ (hypothèse alternative). Notons que la puissance peut en outre s'exprimer en fonction de m_0 et de m_1. Il suffit pour cela de remplacer la valeur critique c par son expression (équation (11.38)) dans l'équation (11.41). Ainsi, il vient :

$$\text{Puissance} = \Phi\left(\frac{m_0 - m_1}{\sigma/\sqrt{n}} + \Phi^{-1}(\alpha)\right) \tag{11.43}$$

Illustrons graphiquement ces notions de risque de première espèce et de risque deuxième espèce (ou de puissance). Pour ce faire, on considère la distribution de la statistique de test, $\overline{X}_n$, obtenue respectivement sous l'hypothèse nulle H_0 et sous l'hypothèse alternative H_1 dans l'exemple précédent. Nous savons que :

$$\overline{X}_n \underset{\mathrm{H}_0}{\sim} \mathcal{N}\left(m_0, \frac{\sigma^2}{\sqrt{n}}\right), \quad \overline{X}_n \underset{\mathrm{H}_1}{\sim} \mathcal{N}\left(m_1, \frac{\sigma^2}{\sqrt{n}}\right) \tag{11.44}$$

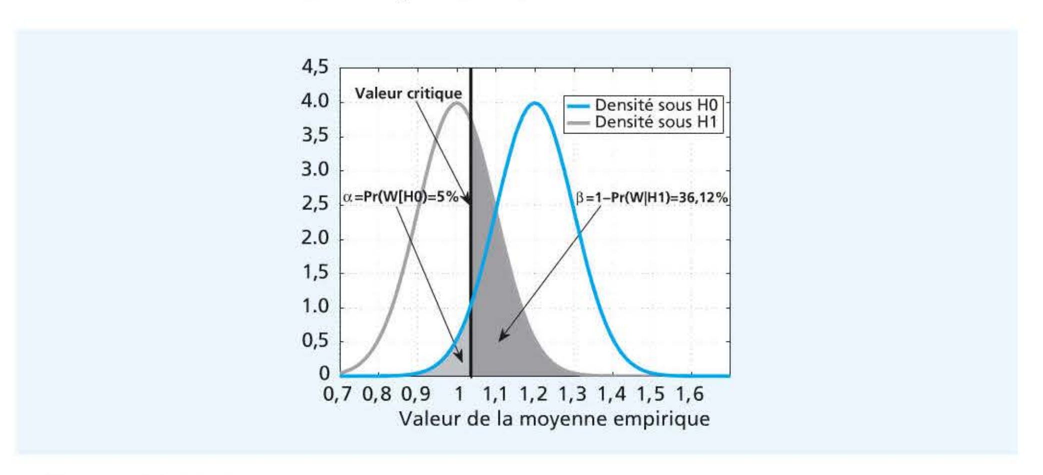
Sur la figure 11.2 sont reportées les fonctions de densité de la variable $\overline{X}_n$ obtenues sous l'hypothèse nulle H_0 et sous l'hypothèse alternative H_1 en fonction

des valeurs de $\overline{x}_n$:

$$\text{Sous } H_0 : \quad f_{\overline{X}_n}(\overline{x}_n; m_0) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\overline{x}_n - m_0}{\sigma/\sqrt{n}}\right)^2\right) \quad \forall \overline{x}_n \in \mathbb{R} \qquad (11.45)$$

$$\text{Sous } H_1 : \quad f_{\overline{X}_n}(\overline{x}_n; m_1) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\overline{x}_n - m_1}{\sigma/\sqrt{n}}\right)^2\right) \quad \forall \overline{x}_n \in \mathbb{R} \qquad (11.46)$$

avec $m_0 = 1{,}2$, $m_1 = 1$, $\sigma^2 = 1$ et $n = 100$. Sous ces hypothèses, nous savons que la région critique du test de niveau $\alpha = 5$ % est définie par W= $\{x : \overline{x}_n < 1{,}0355\}$. La région critique correspond alors à la partie de l'axe des abscisses (valeurs de $\overline{x}_n$) située à gauche de la valeur critique (représentée par une ligne verticale), c'est-à-dire l'ensemble des valeurs $\overline{x}_n$ telles que $\overline{x}_n < 1{,}0355$.



▲ **Figure 11.2 Probabilités de risque I et II**

Comment identifier les probabilités α et β sur ce graphique ? Rappelons que la probabilité α associée à l'erreur de type I, *i.e.* $\alpha = \Pr(W|H_0)$, est définie par la probabilité d'être dans la région critique alors que H_0 est vraie. Elle correspond donc à l'aire sous la densité de $\overline{X}_n$ sous H_0 située à gauche de la valeur critique. La probabilité β associée à l'erreur de type II, *i.e.* $\beta = \Pr\left(\overline{W}\middle|H_1\right)$, est définie par la probabilité de ne pas rejeter H_0, c'est-à-dire de ne pas être dans la région critique, alors que H_1 est vraie. Par conséquent, la probabilité $\beta = 0{,}3612$ correspond à l'aire sous la densité de $\overline{X}_n$ sous l'hypothèse H_1 située à droite de la valeur critique. La puissance, égale à $1-\beta = 0{,}6388$, correspond à l'aire sous la densité sous H_1 située à gauche de la valeur critique.

Nous savons à présent interpréter une région critique et calculer la valeur critique d'un test de niveau α. Mais comment conclure quant à la validité de l'hypothèse nulle ? Il suffit pour cela de vérifier si la réalisation de la statistique de test appartient ou n'appartient pas à la région critique.

Propriété

Règle de décision

Si la réalisation de la statistique de test appartient à la région critique, *on rejette l'hypothèse nulle* H_0 pour un niveau de risque (ou seuil de significativité) α. Si,

au contraire, la réalisation de la statistique de test n'appartient pas à la région critique, on conclut que l'on ne peut pas rejeter l'hypothèse nulle pour un niveau de risque (ou seuil de significativité) α.

Il est donc essentiel de préciser le niveau de risque associé à la décision : on rejette H_0 au seuil de 5 %, de 10 %, etc., car la conclusion peut en effet être tout autre pour un niveau de risque de 15 % par exemple.

Exemple

On considère un n-échantillon de variables $(X_1,\ldots,X_n)$ i.i.d. telles que $X_i \sim \mathcal{N}\left(m,\sigma^2\right)$ avec $\sigma^2 = 1$ et $n = 100$. On souhaite tester :

$$H_0 : m = m_0 = 1{,}2 \quad \text{contre} \quad H_1 : m = m_1 = 1 \tag{11.47}$$

À partir des observations de l'échantillon $(x_1,\ldots,x_n)$, on observe une réalisation de la moyenne empirique égale à $\overline{x}_n = 1{,}13$. Quelle est la conclusion du test pour un seuil de risque $\alpha = 5$ % et un seuil de risque $\alpha = 30$ % ? On admet que la région critique du test de niveau α est définie par :

$$W = \left\{x : \overline{x}_n < m_0 + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(\alpha)\right\} \tag{11.48}$$

Pour $\alpha = 5$ %, on obtient :

$$m_0 + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(\alpha) = 1{,}2 + \frac{1}{\sqrt{100}}\Phi^{-1}(0{,}05) \tag{11.49}$$

$$= 1{,}2 + \frac{1}{\sqrt{100}}(-1{,}6449) = 1{,}0355 \tag{11.50}$$

La région critique du test pour un niveau $\alpha = 5$ % est définie par :

$$W = \{x : \overline{x}_n < 1{,}0355\} \tag{11.51}$$

où $\overline{x}_n$ désigne une réalisation de la statistique de test $\overline{X}_n = n^{-1}\sum_{i=1}^{n} X_i$. Puisque la réalisation de la moyenne empirique, égale à 1,13, n'appartient pas à la région critique on conclut que l'on ne peut pas rejeter l'hypothèse nulle $H_0 : m = 1{,}2$ pour un seuil de significativité de 5 %.

Pour $\alpha = 30$ %, il vient :

$$m_0 + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(\alpha) = 1{,}2 + \frac{1}{\sqrt{100}}\Phi^{-1}(0{,}30) \tag{11.52}$$

$$= 1{,}2 + \frac{1}{\sqrt{100}}(-0{,}5244) = 1{,}1476 \tag{11.53}$$

La région critique du test pour un niveau $\alpha = 30$ % devient :

$$W = \{x : \overline{x}_n < 1{,}1476\} \tag{11.54}$$

Dans ce cas, la réalisation de la moyenne empirique, égale à 1,13, appartient à la région critique. On en conclut que l'on rejette l'hypothèse nulle $H_0 : m = 1{,}2$ pour un seuil de significativité de 30 %. La décision est contraire à celle que nous avions prise pour un niveau $\alpha = 5$ %.

En résumé, la décision issue d'un test peut être :

- soit le rejet de l'hypothèse nulle H_0 pour un *niveau de risque* (ou un seuil de significativité) donné.
- soit le non-rejet de l'hypothèse nulle H_0 pour un *niveau de risque* (ou un seuil de significativité) donné.

FOCUS

La démarche d'un test statistique

À partir des différents éléments présentés jusqu'à présent (hypothèse nulle, hypothèse alternative, région critique, seuil critique, statistique de test, risque de première espèce), nous pouvons résumer la démarche d'un test statistique de la façon suivante :

- **Étape 1.** Poser l'hypothèse nulle H_0 et l'hypothèse alternative H_1, en faisant attention à ce que l'hypothèse nulle corresponde à l'hypothèse pour laquelle le coût associé à l'erreur de type I soit le plus élevé.
- **Étape 2.** Définir la forme de la région critique : cela revient à définir la statistique de test T_n ainsi que la zone de rejet de l'hypothèse nulle H_0 exprimée en fonction des réalisations de cette statistique.
- **Étape 3.** À partir des hypothèses faites sur la (ou les) variable(s) d'intérêt et l'échantillon, dériver la distribution exacte ou la distribution asymptotique de la statistique de test T_n sous l'hypothèse nulle.
- **Étape 4.** Déterminer la (ou les) valeur(s) critique(s) en fonction du niveau de risque α du test.
- **Étape 5.** Calculer la réalisation de la statistique de test T_n à partir des observations de l'échantillon.
- **Étape 6.** Comparer cette réalisation à la région critique du test. Si la réalisation de la statistique de test appartient à la région critique, on conclut au rejet de l'hypothèse nulle H_0 pour un niveau de risque α. Si, au contraire, cette réalisation n'appartient pas à la région critique, on conclut que l'on ne peut pas rejeter l'hypothèse nulle H_0 pour un niveau de risque α.

2.2 La valeur p ou p-value

Pour conclure quant au rejet ou non de l'hypothèse nulle, il convient de comparer la réalisation de la statistique de test à la région critique. Une autre façon de conclure consiste à utiliser la valeur p ou **p-value** en anglais. Nous avons vu, que pour une réalisation donnée de la statistique de test $T_n(x)$ **,** la conclusion du test peut changer lorsque l'on modifie le niveau de risque. L'idée de la p-value consiste à déterminer le plus petit niveau pour lequel on peut rejeter l'hypothèse nulle.

Définition 11.14

Supposons que pour chaque valeur $\alpha \in]0,1[$, corresponde une région critique W_α de niveau α. Alors, la **p-value associée** à une **réalisation de la statistique de test** $T(x)$ est définie comme la plus petite valeur de α pour laquelle on peut rejeter l'hypothèse nulle H_0 :

$$\text{p-value} = \inf \{\alpha : T(x) \in W_\alpha\} \qquad (11.55)$$

Reprenons l'exemple précédent.

Exemple

On considère un n-échantillon de variables $(X_1,\ldots,X_n)$ i.i.d. telles que $X_i \sim \mathcal{N}\left(m,\sigma^2\right)$ avec $\sigma^2 = 1$ et $n = 100$. On souhaite tester :

$$\mathrm{H}_0 : m = m_0 = 1{,}2 \quad \text{contre} \quad \mathrm{H}_1 : m = m_1 = 1 \tag{11.56}$$

À partir des observations de l'échantillon $(x_1,\ldots,x_n)$, on observe une réalisation de la moyenne empirique égale à $\overline{x}_n = 1{,}13$. Déterminons la p-value associée à cette réalisation. On admet que la région critique du test de niveau α est définie par :

$$\mathrm{W}_\alpha = \left\{x : \overline{x}_n < m_0 + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(\alpha)\right\} \tag{11.57}$$

où $\Phi(.)$ désigne la fonction de répartition de la loi normale centrée réduite. Calculons les valeurs critiques pour différentes valeurs de α comprises en 0 et 1. Ces valeurs critiques sont reportées dans le tableau 11.2 avec les conclusions associées quant au rejet ou non de H_0. La p-value associée à $\overline{x}_n = 1{,}13$ correspond à la plus petite valeur de α qui permet de rejeter H_0. Cette p-value est donc comprise entre 0,24 et 0,25.

▼ Tableau 11.2 Valeurs critiques et conclusion du test

α	$\Phi^{-1}(\alpha)$	$m_0 + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(\alpha)$	Conclusion
0,01	−2,3263	0,9674	non-rejet de H_0
0,05	−1,6449	1,0355	non-rejet de H_0
0,10	−1,2816	1,0718	non-rejet de H_0
0,15	−1,0364	1,0964	non-rejet de H_0
0,20	−0,8416	1,1158	non-rejet de H_0
0,24	−0,7063	1,1294	non-rejet de H_0
0,25	−0,6745	1,1326	rejet de H_0
0,30	−0,5244	1,1476	rejet de H_0
0,35	−0,3853	1,1615	rejet de H_0
0,40	−0,2533	1,1747	rejet de H_0

Il existe une façon plus directe de déterminer la p-value pour une réalisation de la statistique de test $T_n(x)$. Pour cela, il suffit de considérer la fonction de répartition de la statistique de test T_n, obtenue sous l'hypothèse nulle à partir, soit de sa loi exacte, soit de sa loi asymptotique. La règle est alors suivante.

Définition 11.15

Suivant la nature du test (unilatéral ou bilatéral), la **p-value** associée à une réalisation $T_n(x)$ est égale à :

$$\text{Test unilatéral droit : p-value} = 1 - F_{T_n}(T_n(x)) \tag{11.58}$$

$$\text{Test unilatéral gauche : p-value} = F_{T_n}(T_n(x)) \tag{11.59}$$

$$\text{Test bilatéral : p-value} = 2 \times F_{T_n}(-|T_n(x)|) \tag{11.60}$$

où $F_{T_n}(.)$ désigne la fonction de répartition de la statistique de test T_n sous l'hypothèse nulle H_0.

Exemple

On considère un n-échantillon de variables $(X_1,\dots,X_n)$ i.i.d. telles que $X_i \sim \mathcal{N}\left(m,\sigma^2\right)$ avec $\sigma^2 = 1$ et $n = 100$. On souhaite tester :

$$\mathrm{H}_0 : m = m_0 = 1{,}2 \quad \text{contre} \quad \mathrm{H}_1 : m = m_1 = 1 \tag{11.61}$$

À partir des observations de l'échantillon $(x_1,\dots,x_n)$, on observe une réalisation de la moyenne empirique égale à $\overline{x}_n = 1{,}13$. Déterminons la p-value associée à cette réalisation. Sous H_0, la statistique de test, *i.e.* la moyenne empirique, admet une distribution (exacte) normale :

$$\overline{X}_n \underset{\mathrm{H}_0}{\sim} \mathcal{N}\left(m_0, \frac{\sigma^2}{n}\right) \Longleftrightarrow \frac{\overline{X}_n - m_0}{\sigma/\sqrt{n}} \underset{\mathrm{H}_0}{\sim} \mathcal{N}(0,1) \tag{11.62}$$

Puisque le test est un test unilatéral gauche, la p-value associée à $\overline{x}_n$ est égale à :

$$\text{p-value} = F_{\overline{X}_n}(\overline{x}_n) = \Pr\left(\overline{X}_n < \overline{x}_n\right) \tag{11.63}$$

On en déduit que :

$$\text{p-value} = \Pr\left(\frac{\overline{X}_n - m_0}{\sigma/\sqrt{n}} < \frac{\overline{x}_n - m_0}{\sigma/\sqrt{n}}\right) = \Phi\left(\frac{\overline{x}_n - m_0}{\sigma/\sqrt{n}}\right) \tag{11.64}$$

où $\Phi(.)$ désigne la fonction de répartition de la loi normale centrée réduite. On vérifie ainsi que la p-value associée à $\overline{x}_n = 1{,}13$ est comprise entre 0,24 et 0,25, puisque :

$$\text{p-value} = \Phi\left(\frac{1{,}13 - 1{,}2}{1/\sqrt{100}}\right) = \Phi(-0{,}7) = 0{,}2420 \tag{11.65}$$

L'avantage de la p-value est qu'elle permet de conclure quant au rejet ou non de H_0 sans avoir à calculer la valeur critique du test. Il suffit de calculer la p-value associée à la réalisation de la statistique de test et d'appliquer la règle de décision suivante.

Propriété

p-value et règle de décision

On rejette l'hypothèse nulle H_0 si la p-value est inférieure au seuil de significativité (niveau) α :

$$\text{p-value} < \alpha \Longrightarrow \text{rejet de } \mathrm{H}_0 \tag{11.66}$$

$$\text{p-value} > \alpha \Longrightarrow \text{non rejet de } \mathrm{H}_0 \tag{11.67}$$

Ainsi, dans l'exemple précédent, nous avions obtenu une p-value de 0,2420. Par conséquent, pour un seuil de significativité $\alpha = 5\ \%$, on conclut au non rejet de l'hypothèse nulle $\mathrm{H}_0 : m = 1{,}2$. Cette conclusion est bien évidemment identique à celle que nous avions obtenue sur la base de la comparaison de la réalisation de la statistique de test et de la région critique définie pour un niveau $\alpha = 5\ \%$.

Remarque : La p-value est donc une mesure du caractère non plausible de l'hypothèse nulle H_0, comme l'indique le tableau 11.3. Mais attention, une p-value importante n'indique pas nécessairement que l'hypothèse nulle H_0 est valide. Dit autrement, la p-value ne correspond pas à la probabilité que H_0 soit vraie. En effet, une p-value peut être importante pour deux raisons : soit parce que H_0 est effectivement vraie, soit parce que le test est peu puissant.

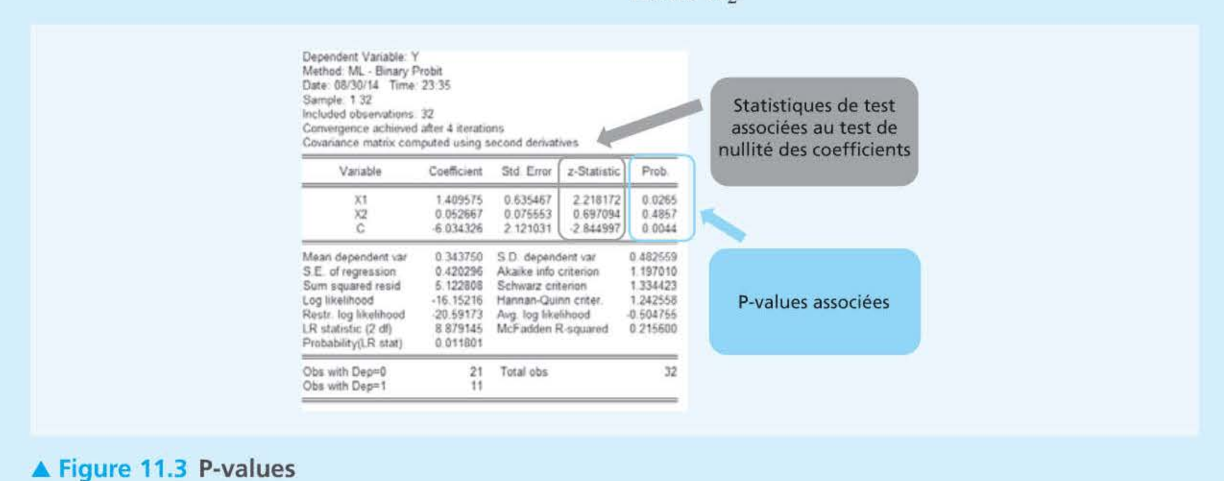
▼ Tableau 11.3 Interprétation de la p-value

p-value	Caractère non plausible de H_0
< 0,01	très forte présomption que H_0 soit fausse
0,01 – 0,05	forte présomption que H_0 soit fausse
0,05 – 0,10	faible présomption que H_0 soit fausse
> 0,10	pas de preuve que H_0 soit fausse

EN PRATIQUE

Les p-values dans les logiciels d'économétrie

Les p-values sont reportées de façon systématique dans la plupart des **logiciels d'économétrie**, comme l'illustre la figure 11.3 reprenant une sortie du logiciel Eviews. Dans cet exemple, l'hypothèse nulle testée est celle de la nullité des coefficients associés aux différentes variables de ce modèle Probit. Sans même connaître le test mis en œuvre, ni sa valeur critique, les p-values permettent de conclure quant au rejet ou non de H_0. C'est pourquoi, elles facilitent grandement l'interprétation des résultats. Ainsi, dans cet exemple, pour un seuil de significativité $\alpha = 5\ \%$, on rejette l'hypothèse nulle de nullité des coefficients associés aux variables X_1 et C (constante). En revanche, on ne peut pas rejeter l'hypothèse nulle de nullité pour le coefficient de la variable X_2.


Dependent Variable: Y
Method: ML - Binary Probit
Date: 08/30/14 Time: 23:35
Sample: 1 32
Included observations: 32
Convergence achieved after 4 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
X1	1.409575	0.635467	2.218172	0.0265
X2	0.052667	0.075553	0.697094	0.4857
C	-6.034326	2.121031	-2.844997	0.0044

Mean dependent var	0.343750	S.D. dependent var	0.482559
S.E. of regression	0.420296	Akaike info criterion	1.197010
Sum squared resid	5.122808	Schwarz criterion	1.334423
Log likelihood	-16.15216	Hannan-Quinn criter.	1.242558
Restr. log likelihood	-20.59173	Avg. log likelihood	-0.504756
LR statistic (2 df)	8.879145	McFadden R-squared	0.215600
Probability(LR stat)	0.011801		
Obs with Dep=0	21	Total obs	32
Obs with Dep=1	11		

▲ Figure 11.3 P-values

2.3 Fonction puissance

Reste une dernière dimension à évoquer concernant la puissance d'un test. Dans le cas d'un test d'une hypothèse simple contre une hypothèse simple, nous avons vu que la puissance était égale à une valeur. Dans le cas d'un test d'une hypothèse simple contre une hypothèse composite, la puissance n'est plus un nombre, mais une fonction de la valeur du paramètre sous H_1 : on parle alors de **fonction puissance**.

Définition 11.16

Lorsque l'hypothèse alternative du test est composite (test unilatéral ou bilatéral), la puissance est une **fonction** de la valeur du paramètre θ sous l'alternative :

$$\text{Puissance} = P(\theta) \quad \forall \theta \in \mathrm{H}_1 \tag{11.68}$$

Intuitivement, plus la valeur de θ sous l'hypothèse alternative H_1 est éloignée de la valeur sous l'hypothèse nulle θ_0, plus la puissance est élevée, car il y a moins de risque de ne pas rejeter H_0 si H_1 est vraie. Inversement, plus la valeur de θ sous l'hypothèse H_1 est proche de θ_0, plus la puissance est faible. Considérons un exemple de fonction puissance.

Exemple

Soit un n-échantillon de variables $(X_1,\ldots,X_n)$ i.i.d., avec $n = 100$, telles que $X_i \sim \mathcal{N}\left(m,\sigma^2\right)$, où m est un paramètre inconnu et $\sigma^2 = 1$. On souhaite tester :

$$\mathrm{H}_0 : m = m_0 = 1{,}2 \quad \text{contre} \quad \mathrm{H}_1 : m < m_0 \tag{11.69}$$

On admet que la région critique du test unilatéral de niveau $\alpha = 5~\%$ est définie par :

$$\mathrm{W} = \left\{x : \overline{x}_n < m_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(\alpha)\right\} \tag{11.70}$$

où $\overline{X}_n = n^{-1} \sum_{i=1}^{n} X_i$ désigne la moyenne empirique. Sous l'hypothèse alternative H_1, nous savons que :

$$\frac{\overline{X}_n - m}{\sigma/\sqrt{n}} \underset{\mathrm{H}_1}{\sim} \mathcal{N}(0,1) \quad \forall m < m_0 \tag{11.71}$$

Par conséquent, la fonction puissance de ce test est définie par :

$$P(m) = \Pr(\mathrm{W}|\,\mathrm{H}_1) = \Pr\left(\overline{X}_n < m_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(\alpha) \middle| H_1\right) \tag{11.72}$$

$$= \Pr\left(\frac{\overline{X}_n - m}{\sigma/\sqrt{n}} < \frac{\sqrt{n}}{\sigma}\left(m_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(\alpha) - m\right) \middle| H_1\right) \tag{11.73}$$

$$= \Phi\left(\frac{m_0 - m}{\sigma/\sqrt{n}} + \Phi^{-1}(\alpha)\right) \quad \forall m < m_0 \tag{11.74}$$

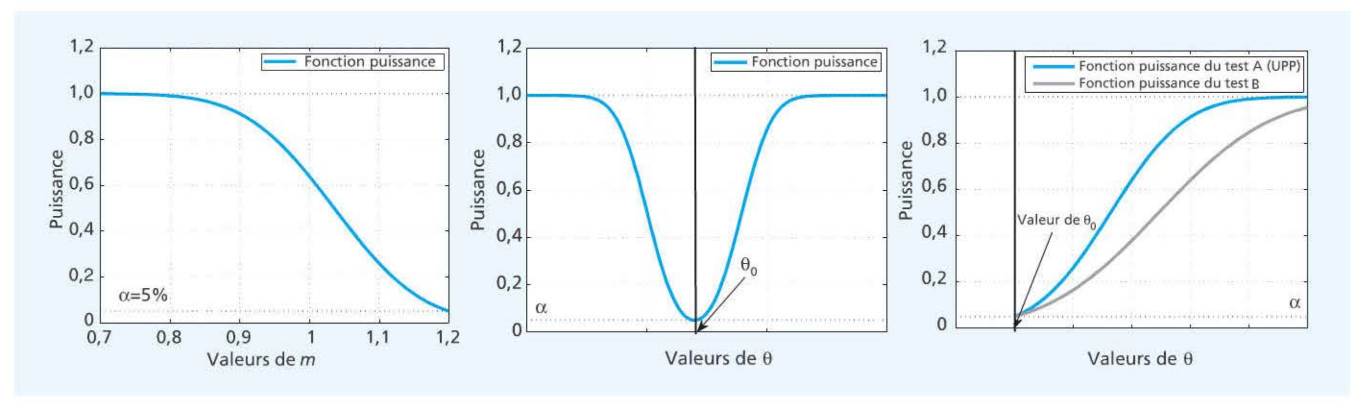
L'application numérique nous donne :

$$P(m) = \Phi\left(\frac{1{,}2 - m}{1/\sqrt{100}} - 1{,}6449\right) \quad \forall m < 1{,}2 \tag{11.75}$$

La figure 11.4 représente cette fonction pour des valeurs de m comprises entre 0,7 et 1,2. On constate que plus la valeur de m sous l'hypothèse alternative s'éloigne de la valeur sous l'hypothèse nulle $m_0 = 1{,}2$, plus la puissance augmente. Pour des valeurs de m suffisamment éloignées de m_0, la puissance est égale à 1. On observe en outre que la puissance du test ne descend jamais en dessous de la taille $\alpha = 5~\%$. Lorsque la valeur de m tend vers la valeur sous l'hypothèse nulle $m_0 = 1{,}2$, la puissance tend vers la taille $\alpha = 5~\%$.

Dans cet exemple, nous avons considéré un test unilatéral du type $\mathrm{H}_0 : \theta = \theta_0$ contre $\mathrm{H}_1 : \theta < \theta_0$. Pour un test unilatéral du type $\mathrm{H}_0 : \theta = \theta_0$ contre $\mathrm{H}_1 : \theta > \theta_0$, la fonction puissance (► figure 11.6) serait définie pour des valeurs $\theta > \theta_0$ sous la forme

d'une fonction croissante avec la valeur de θ. Dans le cas d'un test bilatéral du type $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$, la fonction puissance est définie de part et d'autre de la valeur θ_0 et croit avec la distance $|\theta - \theta_0|$ comme l'illustre la figure 11.5.



▲ **Figure 11.4 Fonction puissance du test unilatéral $H_0 : m = 1,2$ contre $H_1 : m < 1,2$**

▲ **Figure 11.5 Fonction puissance d'un test bilatéral $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$**

▲ **Figure 11.6 Test UPP unilatéral $H_0 : \theta = \theta_0$ contre $H_1 : \theta > \theta_0$**

Remarque : La fonction puissance n'est définie que pour les valeurs de θ admissibles sous l'hypothèse H_1. Ainsi, la quantité $P(\theta_0)$ n'est pas définie.

Nous pouvons à présent définir les notions de **test sans biais** et de **test convergent**.

Définition 11.17

Un test est dit **non biaisé** si la valeur de sa fonction puissance $P(\theta)$ est toujours plus élevée que sa taille α pour toutes les valeurs admissibles de θ sous l'hypothèse alternative H_1 :

$$P(\theta) \geq \alpha \quad \forall \theta \in H_1 \tag{11.76}$$

Par ailleurs, la fonction puissance d'un test non biaisé tend vers la taille lorsque la valeur de θ (sous H_1) tend vers la valeur θ_0 :

$$\lim_{\theta \to \theta_0} P(\theta) = \alpha \tag{11.77}$$

Définition 11.18

Un test est dit **convergent** si sa puissance tend vers l'unité lorsque la taille d'échantillon n tend vers l'infini, quelle que soit la valeur du paramètre θ sous l'hypothèse alternative H_1 :

$$\lim_{n \to \infty} P(\theta) = 1 \quad \forall \theta \in H_1 \tag{11.78}$$

Appliquons ces deux définitions dans le cadre de notre exemple.

Exemple

Soit un n-échantillon de variables $(X_1,\ldots,X_n)$ i.i.d. telles que $X_i \sim \mathcal{N}\left(m,\sigma^2\right)$ où m est un paramètre inconnu. On souhaite tester :

$$H_0 : m = m_0 \quad \text{contre} \quad H_1 : m < m_0 \tag{11.79}$$

On admet que la région critique du test de niveau α est définie par :

$$W = \left\{ x : \overline{x}_n < m_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(\alpha) \right\} \tag{11.80}$$

Nous avons vu que la fonction puissance de ce test est égale à :

$$P(m) = \Phi\left(\frac{m_0 - m}{\sigma / \sqrt{n}} + \Phi^{-1}(\alpha) \right) \quad \forall m < m_0 \tag{11.81}$$

Montrons que ce test est sans biais et convergent. Puisque la quantité $m_0 - m$ est positive, nous avons :

$$\lim_{n\to\infty} \frac{m_0 - m}{\sigma / \sqrt{n}} = +\infty \tag{11.82}$$

La fonction de répartition $\Phi(x)$ étant une fonction croissante à valeurs sur $[0,1]$, on montre que le test est *convergent*, puisque :

$$\lim_{n\to\infty} P(m) = \Phi(+\infty) = 1 \quad \forall m < m_0 \tag{11.83}$$

Par ailleurs, par définition $\Phi\left(\Phi^{-1}(\alpha)\right) = \alpha$, donc :

$$P(m) = \Phi\left(\underbrace{\frac{m_0 - m}{\sigma / \sqrt{n}}}_{\text{quantité positive}} + \Phi^{-1}(\alpha) \right) > \alpha \quad \forall m < m_0 \tag{11.84}$$

La puissance est toujours supérieure à la taille du test, donc le test est sans biais. De plus, on vérifie que :

$$\lim_{m\to m_0} P(m) = \Phi\left(\Phi^{-1}(\alpha)\right) = \alpha \tag{11.85}$$

La comparaison de deux tests convergents et non biaisés, de même niveau se fait sur la base de la fonction puissance. On introduit alors la notion de test uniformément plus puissant ou **test UPP**.

Définition 11.19

Un test A est dit **uniformément plus puissant (UPP)** de niveau α, si sa fonction puissance est supérieure à celle de tous les tests de niveau α pour toutes les valeurs admissibles du paramètre θ sous l'hypothèse alternative H_1 :

$$\alpha_A = \alpha_B = \alpha \qquad P_A(\theta) \geq P_B(\theta) \quad \forall \theta \in H_1 \tag{11.86}$$

pour tout test B de niveau α.

La figure 11.6 illustre, dans le cas d'un test unilatéral du type $H_0 : \theta = \theta_0$ contre $H_1 : \theta > \theta_0$, la notion de test UPP. Sur cette figure sont représentées les fonctions puissance du test UPP de niveau α (test A) et d'un autre test (test B) de même niveau α. La fonction du test UPP est toujours supérieure à celle du test B pour toutes les valeurs de $\theta > \theta_0$. Bien évidemment, lorsque la valeur de θ est très éloignée de θ_0, les puissances des deux tests se rejoignent et tendent vers l'unité.

3 Tests paramétriques

Dans cette section, nous allons nous concentrer sur les **tests paramétriques**, *i.e.* les tests portant sur la valeur d'un ou de plusieurs paramètres de la distribution (paramétrique) de la variable d'intérêt. La question qui se pose est de savoir comment construire un test statistique, c'est-à-dire comment déterminer la forme de la **région critique** de ce test ? Pour ce faire nous allons introduire le **lemme de Neyman-Pearson** dans le cadre de tests d'hypothèses simples, puis nous l'appliquerons à des tests unilatéraux et bilatéraux.

3.1 Lemme de Neyman-Pearson

Le lemme de Neyman-Pearson est une méthode qui permet de dériver la forme de la région critique d'un test paramétrique, c'est-à-dire à la fois la forme de la statistique de test T_n et la forme de la région $\Gamma(c)$.

On considère un n-échantillon $(X_1,...,X_n)$ de variables aléatoires dont la distribution (continue ou discrète) dépend d'un paramètre θ inconnu. Soit $(x_1,...,x_n)$ la réalisation de cet échantillon et $L_n(\theta; x)$ la vraisemblance associée. L'énoncé du lemme de Neyman-Pearson est le suivant :

Propriété

Lemme Neyman-Pearson

Soit le test d'hypothèses simples $\mathrm{H}_0 : \theta = \theta_0$ contre $\mathrm{H}_1 : \theta = \theta_1$. La région critique du test uniformément plus puissant (UPP) de niveau α est définie par :

$$\mathrm{W} = \left\{ x : \frac{L_n(\theta_0; x)}{L_n(\theta_1; x)} < k \right\} \tag{11.87}$$

où $L_n(\theta; x)$ désigne la vraisemblance de l'échantillon $(x_1,...,x_n)$ et k est une constante déterminée par le niveau du test α, telle que :

$$\Pr\left(\frac{L_n(\theta_0; X)}{L_n(\theta_1; X)} < k \middle| \mathrm{H}_0 \right) = \alpha \tag{11.88}$$

Comment utiliser le lemme de Neyman-Pearson ? L'idée est de réarranger les termes de l'inégalité $L_n(\theta_0; x)/L_n(\theta_1; x) < k$ afin d'obtenir un résultat du type :

$$T_n(x) \leqslant c \tag{11.89}$$

où c est une valeur critique (constante) déterminée par le niveau α du test et $T_n(x)$ est une réalisation de la statistique de test T_n dont on connaît la loi exacte ou la loi asymptotique sous l'hypothèse nulle H_0. Pour cela, on doit donc rassembler à gauche de l'inégalité les termes dépendant des réalisations $x_1,...,x_n$, et à droite les termes constants. Il convient toutefois de bien faire attention au sens de l'inégalité qui peut changer. Considérons un exemple d'application du lemme de Neyman-Pearson.

Exemple

On considère un n-échantillon $(X_1,\ldots,X_n)$, avec $n = 100$, de variables i.i.d. telles que $X_i \sim \mathcal{N}\left(m,\sigma^2\right)$ où le paramètre m est inconnu et $\sigma^2 = 1$. On souhaite tester :

$$\text{H}_0 : m = m_0 = 1{,}2 \quad \text{contre} \quad \text{H}_1 : m = m_1 = 1{,}4 \tag{11.90}$$

Quelle est la région critique du test UPP de niveau $\alpha = 5\ \%$? Puisque les variables $X_1,\ldots,X_n$ sont N.i.d.$\left(m,\sigma^2\right)$, la vraisemblance de l'échantillon $(x_1,\ldots,x_n)$ est définie par (► chapitre 10) :

$$L_n(m; x) = \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2\right) \tag{11.91}$$

D'après le lemme de Neyman-Pearson, la région critique du test UPP de niveau α est de la forme :

$$\frac{L_n(m_0; x)}{L_n(m_1; x)} < k \tag{11.92}$$

où k est une constante déterminée par le niveau α. En utilisant la forme de la vraisemblance sous H_0 et H_1, il vient :

$$\frac{L_n(m_0; x)}{L_n(m_1; x)} = \frac{\frac{1}{\sigma^n(2\pi)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m_0)^2\right)}{\frac{1}{\sigma^n(2\pi)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m_1)^2\right)} < k \tag{11.93}$$

Réarrangeons ces termes de sorte à isoler à gauche une statistique de test et à droite un terme constant. Cette inégalité peut se réécrire sous la forme :

$$\exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i - m_0)^2 - \sum_{i=1}^{n}(x_i - m_1)^2\right)\right) < k \tag{11.94}$$

$$\Longleftrightarrow \sum_{i=1}^{n}(x_i - m_1)^2 - \sum_{i=1}^{n}(x_i - m_0)^2 < k_1 \tag{11.95}$$

où $k_1 = 2\sigma^2 \ln(k)$ est une constante. Ainsi, nous avons :

$$2(m_0 - m_1)\sum_{i=1}^{n} x_i + n\left(m_1^2 - m_0^2\right) < k_1 \Longleftrightarrow (m_0 - m_1)\sum_{i=1}^{n} x_i < k_2 \tag{11.96}$$

où $k_2 = \left(k_1 - n\left(m_1^2 - m_0^2\right)\right)/2$ est une constante. Puisque $m_1 - m_0 = 0{,}2 > 0$, nous obtenons finalement :

$$\frac{1}{n}\sum_{i=1}^{n} x_i > k_3 \tag{11.97}$$

où $k_3 = k_2/(n(m_0 - m_1))$ est une constante. Par conséquent, la région critique du test UPP a une forme générale du type :

$$\text{W} = \{x : \overline{x}_n > c\} \tag{11.98}$$

où c est constante (valeur critique) déterminée par le niveau α. La statistique de test correspond à la moyenne empirique $\overline{X}_n = \sum_{i=1}^{n} X_i$ et vérifie :

$$\frac{\overline{X}_n - m_0}{\sigma/\sqrt{n}} \underset{\text{H}_0}{\sim} \mathcal{N}(0,1) \tag{11.99}$$

On remarque que la forme des constantes k_1, k_2 et k_3 n'a aucune importance. Ce qui importe c'est que ces paramètres ne dépendent pas des réalisations $x_1,\ldots x_n$. Comme nous l'avons fait dans la section 1.3, nous pouvons exprimer la valeur critique A en fonction de α :

$$\alpha = \Pr(\text{W}|\,\text{H}_0) = \Pr(\overline{x}_n > c|\,\text{H}_0) \tag{11.100}$$

$$= 1 - \Pr\left(\frac{\overline{X}_n - m_0}{\sigma/\sqrt{n}} < \frac{c - m_0}{\sigma/\sqrt{n}}\middle|\,\text{H}_0\right) \tag{11.101}$$

$$= 1 - \Phi\left(\frac{c - m_0}{\sigma/\sqrt{n}}\right) \tag{11.102}$$

où $\Phi(.)$ désigne la fonction de répartition de la loi normale centrée réduite.

$$1-\alpha=\Phi\left(\frac{c-m_0}{\sigma/\sqrt{n}}\right)\Longleftrightarrow\Phi^{-1}(1-\alpha)=\frac{c-m_0}{\sigma/\sqrt{n}} \quad (11.103)$$

On en déduit la valeur critique du test :

$$c=m_0+\frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha)=1{,}2+\frac{1}{\sqrt{100}}\times\Phi^{-1}(0{,}95) \quad (11.104)$$

$$=1{,}2+\frac{1}{\sqrt{100}}\times 1{,}6449=1{,}3645 \quad (11.105)$$

Au final, la région critique du test UPP de niveau $\alpha = 5\ \%$ de l'hypothèse $H_0 : m = m_0 = 1{,}2$ contre $H_1 : m = m_1 = 1{,}4$, est définie par :

$$W=\left\{x:\overline{x}_n>m_0+\frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha)\right\}=\{x:\overline{x}_n>1{,}3645\} \quad (11.106)$$

Si la réalisation de la moyenne empirique est supérieure à 1,3645, on rejette l'hypothèse nulle $H_0 : m = 1{,}2$ pour un seuil de significativité de 5 %.

Remarque : Dans cet exemple, la statistique de test correspond à un estimateur du paramètre testé, puisque $\overline{X}_n$ est un estimateur de l'espérance $\mathbb{E}(X_i) = m$. Dans ce cas, la forme de la région critique est évidente. Puisque la réalisation de l'estimateur est censée être « proche » de la vraie valeur, si l'on teste :

$$H_0:\theta=\theta_0 \text{ contre } H_1:\theta=\theta_1 \text{ avec } \theta_1>\theta_0, \quad (11.107)$$

on rejette l'hypothèse nulle H_0 lorsque la réalisation de l'estimateur est suffisamment « grande », c'est-à-dire lorsque cette réalisation est supérieure à une certaine valeur critique. La région critique du test UPP est donc de la forme :

$$W=\left\{x:\widehat{\theta}(x)>c\right\} \quad (11.108)$$

où $\widehat{\theta}(x)$ désigne la réalisation de l'estimateur $\widehat{\theta}$. Inversement, si l'on teste :

$$H_0:\theta=\theta_0 \text{ contre } H_1:\theta=\theta_1 \text{ avec } \theta_1<\theta_0, \quad (11.109)$$

on rejette l'hypothèse nulle H_0 lorsque la réalisation de l'estimateur est suffisamment « petite », et la région critique est de la forme :

$$W=\left\{x:\widehat{\theta}(x)<c\right\} \quad (11.110)$$

3.2 Tests unilatéraux et bilatéraux

Nous savons à présent comment déterminer la région critique du test UPP d'une hypothèse simple contre une hypothèse simple. Mais qu'en est-il pour les tests unilatéraux et bilatéraux ?

Définition 11.20

La région critique du **test unilatéral** UPP de niveau α :

$$H_0:\theta=\theta_0 \text{ contre } H_1:\theta>\theta_0 \text{ (ou } H_1:\theta<\theta_1) \quad (11.111)$$

est équivalente à celle du test d'hypothèses simples :

$$H_0:\theta=\theta_0 \text{ contre } H_1:\theta=\theta_1 \quad (11.112)$$

avec $\theta_1 > \theta_0$ (ou $\theta_1 < \theta_0$), dès lors que cette région ne dépend pas de la valeur de θ_1.

Appliquons cette définition dans le cadre de notre exemple.

Exemple

On considère un n-échantillon $(X_1,\ldots,X_n)$ de variables N.i.d.$\left(m,\sigma^2\right)$ où le paramètre m est inconnu. On souhaite tester :

$$\mathrm{H}_0 : m = m_0 \quad \text{contre} \quad \mathrm{H}_0 : m > m_0 \tag{11.113}$$

Déterminons la région critique du test UPP de taille α. Pour cela, on considère le test d'hypothèses simples :

$$\mathrm{H}_0 : m = m_0 \quad \text{contre} \quad \mathrm{H}_0 : m = m_1 \tag{11.114}$$

où m_1 est une valeur telle que $m_1 > m_0$. D'après le lemme de Neyman-Pearson, la région critique du test UPP de niveau α est (▸ exemple précédent) :

$$\mathrm{W} = \left\{ x : \overline{x}_n > m_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1-\alpha) \right\} \tag{11.115}$$

où $\Phi(.)$ désigne la fonction de répartition de la loi normale centrée réduite. La région W ne dépend pas du choix de la valeur de m_1. Cette région correspond donc à celle du test UPP unilatéral de niveau α :

$$\mathrm{H}_0 : m = m_0 \quad \text{contre} \quad \mathrm{H}_0 : m > m_0 \tag{11.116}$$

Dans le cas d'un **test bilatéral** $\mathrm{H}_0 : \theta = \theta_0$ contre $\mathrm{H}_1 : \theta \neq \theta_0$, il n'existe pas de test UPP valable à la fois pour les valeurs de θ supérieures à la valeur nulle θ_0 et pour les valeurs inférieures à ce seuil. Dit autrement, si l'on considère deux tests A et B tels que la fonction puissance du test A est supérieure à celle du test B pour les valeurs $\theta < \theta_0$, alors la fonction puissance du test A est nécessairement inférieure à celle du test B pour les valeurs $\theta > \theta_0$. C'est pourquoi, la région de non-rejet du test bilatéral (non UPP) est définie par l'intersection des régions de non rejet des tests unilatéraux UPP correspondants.

Définition 11.21

La région de non rejet $\overline{\mathrm{W}}$ du **test bilatéral** de niveau α :

$$\mathrm{H}_0 : \theta = \theta_0 \ \text{contre}\ \mathrm{H}_1 : \theta \neq \theta_0 \tag{11.117}$$

est définie par l'intersection des régions de non rejet des tests unilatéraux UPP correspondants de niveau $\alpha/2$:

$$\text{Test A} : \ \mathrm{H}_0 : \theta = \theta_0 \ \text{contre}\ \mathrm{H}_1 : \theta > \theta_0 \tag{11.118}$$

$$\text{Test B} : \ \mathrm{H}_0 : \theta = \theta_0 \ \text{contre}\ \mathrm{H}_1 : \theta < \theta_0 \tag{11.119}$$

Soient $\overline{\mathrm{W}}_A$ et $\overline{\mathrm{W}}_B$ les régions de non rejet des tests A et B au niveau $\alpha/2$, la région critique du test bilatéral de niveau α vérifie :

$$\overline{\mathrm{W}} = \overline{\mathrm{W}}_A \cap \overline{\mathrm{W}}_B \tag{11.120}$$

Il est important de noter que les seuils critiques des tests unilatéraux qui servent à construire la région critique du test bilatéral de niveau α doivent être considérés pour un niveau de risque $\alpha/2$ et non α.

FOCUS

La région critique d'un test bilatéral

La région de non rejet du test bilatéral $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ de niveau α, est définie par l'intersection des régions de non-rejet des tests unilatéraux UPP associés de niveau $\alpha/2$. Pourquoi utiliser un niveau de risque $\alpha/2$ au lieu de α ? Raisonnons par l'absurde. Supposons que les régions critiques et les régions de non rejet des tests unilatéraux s'écrivent sous la forme :

$$\text{Test A} : H_0 : \theta = \theta_0 \quad \text{contre} \quad H_1 : \theta > \theta_0 \tag{11.121}$$

$$\text{Test B} : H_0 : \theta = \theta_0 \quad \text{contre} \quad H_1 : \theta < \theta_0 \tag{11.122}$$

$$W_A = \{x : T_n(x) > c_A\} \qquad \overline{W}_A = \{x : T_n(x) < c_A\} \tag{11.123}$$

$$W_B = \{x : T_n(x) < c_B\} \qquad \overline{W}_B = \{x : T_n(x) > c_B\} \tag{11.124}$$

où $T_n(x)$ désigne la réalisation de la statistique de test T_n, et c_A et c_B sont deux valeurs critiques. Si l'on suppose que ces régions critiques sont construites pour un niveau de risque de première espèce égal à α, alors par définition :

$$\alpha = \Pr(W_A | H_0) = \Pr(T_n > c_A | H_0) \tag{11.125}$$

$$\alpha = \Pr(W_B | H_0) = \Pr(T_n < c_B | H_0) \tag{11.126}$$

Logiquement, les valeurs critiques c_A et c_B vérifient $c_B < c_A$. Dès lors, la région de non-rejet $\overline{W}$ du test bilatéral, définie par l'intersection des régions $\overline{W}_A$ et $\overline{W}_B$, est égale à :

$$\overline{W} = \overline{W}_A \cap \overline{W}_B = \{x : (T_n(x) < c_A) \cap (T_n(x) > c_B)\} \tag{11.127}$$

On obtient ainsi :

$$\overline{W} = \{x : c_B < T_n(x) < c_A\} \tag{11.128}$$

Déterminons le risque de première espèce associé à ce test bilatéral :

$$\begin{aligned} \Pr(W | H_0) &= 1 - \Pr\left(\overline{W} \middle| H_0\right) \\ &= 1 - \Pr(c_B < T_n < c_A | H_0) \end{aligned} \tag{11.129}$$

Sachant que $\Pr(u < X < v) = \Pr(X < v) - \Pr(X < u)$, il vient :

$$\begin{aligned} \Pr(W | H_0) = 1 &- \Pr(T_n < c_A | H_0) \\ &+ \Pr(T_n < c_B | H_0) \end{aligned} \tag{11.130}$$

$$= 1 - (1 - \alpha) + \alpha = 2\alpha \tag{11.131}$$

Le niveau de risque du test bilatéral est donc égal à 2α. C'est pourquoi, afin d'obtenir un niveau de risque précisément égal à α pour le test bilatéral, on considère des seuils critiques des tests unilatéraux définis pour un niveau de risque de $\alpha/2$.

Appliquons cette définition dans le cadre de notre exemple.

Exemple

On considère un n-échantillon $(X_1, \ldots, X_n)$, avec $n = 100$, de variables N.i.d$\left(m, \sigma^2\right)$, où m est un paramètre inconnu et $\sigma^2 = 1$. On souhaite tester :

$$H_0 : m = m_0 = 1{,}2 \quad \text{contre} \quad H_1 : m \neq m_0 \tag{11.132}$$

Déterminons la région critique de ce test bilatéral pour un niveau $\alpha = 5\ \%$. On considère les tests unilatéraux associés :

$$\text{Test A} : \quad H_0 : m = m_0 \quad \text{contre} \quad H_1 : m < m_0 \tag{11.133}$$

$$\text{Test B} : \quad H_0 : m = m_0 \quad \text{contre} \quad H_1 : m > m_0 \tag{11.134}$$

Les régions critiques des tests UPP de niveau $\alpha/2$ sont définies par :

$$W_A = \left\{x : \overline{x}_n < m_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(\frac{\alpha}{2}\right)\right\} \tag{11.135}$$

$$W_B = \left\{x : \overline{x}_n > m_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right\} \tag{11.136}$$

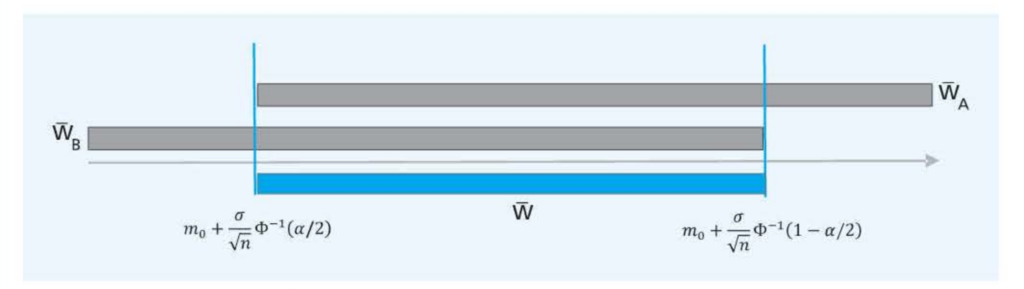
où $\Phi(.)$ désigne la fonction de répartition de la loi normale centrée réduite. Les régions de non rejet de niveau $\alpha/2$ sont définies comme les régions complémentaires de W_A et de W_B :

$$\overline{W}_A = \left\{ x : \overline{x}_n \geq m_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(\frac{\alpha}{2}\right) \right\} \tag{11.137}$$

$$\overline{W}_B = \left\{ x : \overline{x}_n \leq m_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \right\} \tag{11.138}$$

La région de non rejet du test bilatéral de niveau α correspond à la zone d'intersection de ces deux régions :

$$\overline{W} = \overline{W}_A \cap \overline{W}_B \tag{11.139}$$



▲ **Figure 11.7 Région de non-rejet du test bilatéral de niveau α**

Pour $\alpha = 5\ \%$, nous savons que $\Phi^{-1}(\alpha/2) < 0$ et $\Phi^{-1}(1-\alpha/2) > 0$. Par conséquent, les valeurs critiques des deux tests unilatéraux vérifient :

$$m_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(\frac{\alpha}{2}\right) < m_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \tag{11.140}$$

Comme le montre la figure 11.7, la *région de non-rejet* du test bilatéral de niveau α est donc définie par :

$$\overline{W} = \left\{ x : m_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(\frac{\alpha}{2}\right) \leq \overline{x}_n \leq m_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \right\} \tag{11.141}$$

Sachant que $n = 100$, $m_0 = 1{,}2$, $\sigma^2 = 1$ et $\alpha = 5\ \%$, nous avons :

$$m_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(\frac{\alpha}{2}\right) = 1{,}2 + \frac{1}{\sqrt{100}} \times (-1{,}96) = 1{,}0040 \tag{11.142}$$

$$m_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 1{,}2 + \frac{1}{\sqrt{100}} \times (1{,}96) = 1{,}3960 \tag{11.143}$$

La région de non-rejet et la région critique du test bilatéral de niveau $\alpha = 5\ \%$ sont respectivement définies par :

$$\overline{W} = \{x : 1{,}0040 \leq \overline{x}_n \leq 1{,}3960\}, \quad W = \{x : \overline{x}_n \notin [1{,}0040\ ; 1{,}3960]\} \tag{11.144}$$

On peut réécrire ces deux régions sous une autre forme. Puisque la loi normale standard est symétrique par rapport à zéro, on a $\Phi^{-1}(\alpha/2) = -\Phi^{-1}(1-\alpha/2)$. La région de non-rejet devient :

$$\overline{W} = \left\{ x : m_0 - \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \leq \overline{x}_n \leq m_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \right\} \tag{11.145}$$

ou encore :

$$\overline{W} = \left\{ x : \left| \frac{\overline{x}_n - m_0}{\sigma/\sqrt{n}} \right| \leq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \right\} \tag{11.146}$$

La région critique du test bilatéral de niveau α peut donc être définie sous la forme :

$$W = \left\{ x : \left| \frac{\overline{x}_n - m_0}{\sigma/\sqrt{n}} \right| > \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \right\} \tag{11.147}$$

Soit W= $\left\{x : |\overline{x}_n - 1{,}2| / \left(\sigma/\sqrt{n}\right) > 1{,}96\right\}$. Si l'écart entre la réalisation de la moyenne empirique et la valeur nulle $m_0 = 1{,}2$, normalisé par $\sigma/\sqrt{n}$, est supérieur (en valeur absolue) au seuil critique de 1,96, on rejette l'hypothèse nulle $H_0 : m = 1{,}2$ pour un niveau de risque de 5 %.

4 Tests d'indépendance et d'adéquation

Jusqu'à présent, nous avons considéré des tests paramétriques portant sur la valeur d'un paramètre de la distribution de la variable d'intérêt dans la population. Dans cette section nous allons étudier deux tests non-paramétriques : le **test d'indépendance** du khi-deux et le **test d'adéquation** (ou d'ajustement) du khi-deux. Ces tests sont particulièrement utilisés, notamment dans le cadre d'applications en marketing.

4.1 Tests d'indépendance du khi-deux

Définition 11.22

Le **test d'indépendance du khi-deux** permet de tester si deux variables aléatoires, X et Y, sont indépendantes. Les hypothèses nulle et alternative de ce test s'écrivent respectivement sous la forme :

$$H_0 : X \text{ et } Y \text{ sont indépendantes contre } H_1 : X \text{ et } Y \text{ sont dépendantes} \quad (11.148)$$

Nous considérerons ici le cas de deux variables discrètes[2] admettant un nombre de modalités fini. On suppose que la variable aléatoire discrète X peut prendre *k modalités différentes* et que la variable Y peut prendre *s modalités différentes* :

$$X = \begin{cases} a_1 & \text{avec une probabilité } \Pr(X = a_1) \\ a_2 & \text{avec une probabilité } \Pr(X = a_2) \\ \dots & .. \\ a_k & \text{avec une probabilité } \Pr(X = a_k) \end{cases} \quad (11.149)$$

$$Y = \begin{cases} b_1 & \text{avec une probabilité } \Pr(Y = b_1) \\ b_2 & \text{avec une probabilité } \Pr(Y = b_2) \\ \dots & .. \\ b_s & \text{avec une probabilité } \Pr(Y = b_s) \end{cases} \quad (11.150)$$

avec par définition $\sum_{i=1}^{k} \Pr(X = a_i) = 1$ et $\sum_{j=1}^{s} \Pr\left(Y = b_j\right) = 1$.

Pour mettre en œuvre ce test, on dipose d'un n-échantillon $(x_v, y_v)_{v=1}^{n}$. Ces observations peuvent être représentées par un **tableau de contingence** (▶ chapitre 2). Soit $n_{i,j}$ le

2 Le test d'indépendance du khi-deux peut aussi s'appliquer à des variables continues ou à des variables discrètes admettant un nombre de modalités infini (par exemple si $X \in \mathbb{N}$). Dans ce cas, on découpe l'ensemble des valeurs que peut prendre X en k classes et l'on fait de même pour Y en découpant son support en s classes.

nombre d'individus dans l'échantillon pour lesquels on observe $X = a_i$ et $Y = b_j$. Soit $n_{x=i}$ le nombre total d'individus pour lesquels on observe $X = a_i$ et soit $n_{y=j}$ le nombre d'individus pour lesquels on observe $Y = b_j$, tels que :

$$n_{x=i} = \sum_{v=1}^{n} I_{(X_v=a_i)} = \sum_{j=1}^{s} n_{i,j}, \quad n_{y=j} = \sum_{v=1}^{n} I_{(Y_v=b_j)} = \sum_{i=1}^{k} n_{i,j} \tag{11.151}$$

où la fonction $I(.)$ est égale à 1 si la condition $(.)$ est vérifiée et à 0 sinon.

L'idée du test d'indépendance du Khi-deux est de comparer le *tableau de contingence empirique* 11.4 (basé sur les observations de l'échantillon) à un *tableau de contingence théorique* que l'on obtiendrait si les deux variables X et Y étaient effectivement indépendantes. Si les deux tableaux sont similaires, alors on ne peut pas rejeter l'hypothèse nulle H_0 d'indépendance. En revanche, si ces deux tableaux sont suffisamment différents, on rejette l'hypothèse nulle H_0 d'indépendance.

▼ Tableau 11.4 Tableau de contingence empirique

$X \setminus Y$	$Y = b_1$	..	$Y = b_j$	..	$Y = b_s$	Total
$X = a_1$	$n_{1,1}$	..	$n_{1,j}$	..	$n_{1,s}$	$n_{x=1}$
..	..	..	..	..	..	..
$X = a_i$	$n_{i,1}$	..	$n_{i,j}$	..	$n_{i,s}$	$n_{x=i}$
..	..	..	..	..	..	..
$X = a_k$	$n_{k,1}$	..	$n_{k,j}$	..	$n_{k,s}$	$n_{x=k}$
Total	$n_{y=1}$	..	$n_{y=j}$	..	$n_{y=s}$	n

Comment construire le tableau de contingence théorique sous l'hypothèse nulle d'indépendance ? Pour cela, il faut connaître le nombre théorique, noté $N_{i,j}$, d'individus pour lesquels on devrait observer à la fois $X = a_i$ et $Y = b_j$ si les variables X et Y étaient effectivement indépendantes. Par définition, cet effectif théorique est égal à :

$$N_{i,j} = n \times \Pr\left((X = a_i) \cap \left(Y = b_j\right)\right) \tag{11.152}$$

Or, sous l'hypothèse d'indépendance, nous savons que la probabilité jointe d'observer $X = a_i$ et $Y = b_j$ est égale au produit des probabilités marginales (► chapitre 6) :

$$N_{i,j} = n \times \Pr(X = a_i) \times \Pr\left(Y = b_j\right) \tag{11.153}$$

Les probabilités marginales $\Pr(X = a_i)$ et $\Pr\left(Y = b_j\right)$ étant inconnues, il convient de les estimer. Sachant qu'un estimateur convergent de la probabilité associée à un événement est donné par la fréquence empirique d'apparition de cet événement, on peut estimer $\Pr(X = a_i)$ et $\Pr\left(Y = b_j\right)$ de la façon suivante :

$$\widehat{\Pr}(X = a_i) = \frac{n_{x=i}}{n} \qquad \widehat{\Pr}\left(Y = b_j\right) = \frac{n_{y=j}}{n} \tag{11.154}$$

Ainsi, l'estimateur des effectifs théoriques $N_{i,j}$ devient :

$$\widehat{N}_{i,j} = n \times \widehat{\Pr}(X = a_i) \times \widehat{\Pr}\left(Y = b_j\right) = n \times \frac{n_{x=i}}{n} \times \frac{n_{y=j}}{n} = \frac{n_{x=i} \times n_{y=j}}{n} \tag{11.155}$$

Définition 11.23

Un **estimateur des effectifs théoriques** $N_{i,j}$ pour $i = 1,\ldots,k$ et $j = 1,\ldots,s$ est défini par :

$$\widehat{N}_{i,j} = \frac{n_{x=i} \times n_{y=j}}{n} \quad (11.156)$$

où $n_{x=i}$ et $n_{y=j}$ désignent respectivement le nombre d'individus dans l'échantillon pour lesquels $X = a_i$ et $Y = b_j$, et n correspond à la taille de l'échantillon.

On obtient ainsi un tableau de contingence théorique, comme reproduit dans le tableau 11.5.

▼ Tableau 11.5 Tableau de contingence théorique

$X \setminus Y$	$Y = b_1$	..	$Y = b_j$	..	$Y = b_s$	Total
$X = a_1$	$\widehat{N}_{1,1} = \frac{n_{x=1} \times n_{y=1}}{n}$	..	$\widehat{N}_{1,j} = \frac{n_{x=1} \times n_{y=j}}{n}$	..	$\widehat{N}_{1,s} = \frac{n_{x=1} \times n_{y=s}}{n}$	$n_{x=1}$
..	..	..	..	..	..	..
$X = a_i$	$\widehat{N}_{i,1} = \frac{n_{x=i} \times n_{y=1}}{n}$	..	$\widehat{N}_{i,j} = \frac{n_{x=i} \times n_{y=j}}{n}$	..	$\widehat{N}_{i,s} = \frac{n_{x=i} \times n_{y=s}}{n}$	$n_{x=i}$
..	..	..	..	..	..	..
$X = a_k$	$\widehat{N}_{k,1} = \frac{n_{x=k} \times n_{y=1}}{n}$	..	$\widehat{N}_{k,j} = \frac{n_{x=k} \times n_{y=j}}{n}$	..	$\widehat{N}_{k,s} = \frac{n_{x=k} \times n_{y=s}}{n}$	$n_{x=k}$
Total	$n_{y=1}$	..	$n_{y=j}$	..	$n_{y=s}$	n

Remarque : Par construction, les sommes des colonnes et des lignes du tableau de contingence théorique correspondent à celles du tableau de contingence empirique : $\sum_{j=1}^{s} \widehat{N}_{i,j} = n_{x=i}$ et $\sum_{i=1}^{k} \widehat{N}_{i,j} = n_{y=j}$.

Comment comparer ces deux tableaux de contingence et conclure quant au rejet ou non de l'hypothèse nulle H_0 ? On utilise pour cela le **test d'indépendance du khi-deux** dont la région critique et la statistique de test sont définies de la façon suivante.

Définition 11.24

La **statistique de test d'indépendance du khi-deux**, notée D_n, est définie par :

$$D_n = \sum_{i=1}^{k} \sum_{j=1}^{s} \frac{(n_{i,j} - \widehat{N}_{i,j})^2}{\widehat{N}_{i,j}} \quad (11.157)$$

Sous l'hypothèse nulle d'indépendance, la statistique de test D_n admet une distribution exacte du khi-deux à $(k-1) \times (s-1)$ degrés de liberté :

$$D_n \underset{H_0}{\sim} \chi^2((k-1) \times (s-1)) \quad (11.158)$$

L'idée de la statistique du khi-deux est de mesurer la distance entre les effectifs empiriques $n_{i,j}$ et les effectifs théoriques estimés $\widehat{N}_{i,j}$ pour les $k \times s$ cases des tableaux de contingence, c'est-à-dire pour $i = 1,\ldots,k$ et $j = 1,\ldots,s$. Le carré permet que les écarts

négatifs ne compensent pas des écarts positifs. La normalisation par $\widehat{N}_{i,j}$ permet que la statistique de test ne diverge pas. Sous H_0, cette statistique suit une loi du khi-deux à $(k-1)\times(s-1)$ degrés de liberté, car bien évidemment les effectifs théoriques ne sont pas indépendants, puisque leur somme sur X ou sur Y est égale à la taille d'échantillon n. C'est pour cela que l'on ajuste les degrés de liberté en enlevant une unité aux dimensions k et s. La région critique du test est alors la suivante.

Définition 11.25

La **région critique** du test d'indépendance du khi-deux pour un niveau de risque α est définie par :

$$W = \left\{x,y : D_n(x,y) > G_d^{-1}(1-\alpha)\right\} \quad (11.159)$$

où $D_n(x,y)$ désigne la réalisation de la statistique D_n et $G_d(.)$ est la fonction de répartition de la loi du khi-deux à $d = (k-1)\times(s-1)$ degrés de liberté.

Ainsi, si la réalisation de la statistique du khi-deux est supérieure au fractile à $100\times(1-\alpha)$ % de la loi du khi-deux à $(k-1)\times(s-1)$ degrés de liberté, on rejette l'hypothèse nulle d'indépendance pour un seuil de risque α. En effet, cela signifie que la statistique du khi-deux est trop importante par rapport au seuil critique, c'est-à-dire que la distance entre les effectifs empiriques et les effectifs théoriques est elle-même trop grande sur, au moins, l'une des $k\times s$ classes. On doit donc rejeter H_0. Appliquons ce test dans le cadre d'un exemple d'étude marketing.

Exemple

Une entreprise souhaite analyser l'impact d'une campagne marketing suivant les trois canaux de diffusion utilisés : emailing, courriers et appels téléphoniques. Pour cela, elle dispose d'un échantillon de 1 500 clients ayant été contactés par l'un des trois médias, représenté par le tableau de contingence 11.6.

▼ **Tableau 11.6 Tableau de contingence empirique**

Montant \ Média	Emailing	Courriers	Appels	Total
50-100 euros	220	200	50	470
100-200 euros	140	250	100	490
plus de 200 euros	140	50	350	540
Total	**500**	**500**	**500**	**1 500**

Ce tableau s'interprète de la façon suivante : par exemple, 220 clients contactés par mail ont acheté en moyenne entre 50 et 100 euros de produits. On souhaite tester l'indépendance entre le montant moyen des achats (variable X) et le média (variable Y) au seuil de risque de 10 %. Pour ce faire, construisons le tableau de contingence théorique obtenu sous l'hypothèse nulle d'indépendance à partir des effectifs théoriques estimés :

$$\widehat{N}_{i,j} = \frac{n_{x=i}\times n_{y=j}}{n} \quad (11.160)$$

Par exemple, l'effectif théorique des consommateurs ayant été contactés par emailing et ayant consommé entre 50 et 100 euros est égal à :

$$\widehat{N}_{x=50/100,email} = \frac{n_{x=50/100}\times n_{y=email}}{n} = \frac{470\times 500}{1\,500} = 156,66 \quad (11.161)$$

De la même façon, l'effectif théorique des consommateurs ayant été contactés par email et ayant consommé entre 100 et 200 euros est égal à :

$$\widehat{N}_{x=100/200,email} = \frac{n_{x=100/200} \times n_{y=email}}{n} = \frac{490 \times 500}{1\,500} = 163{,}33 \qquad (11.162)$$

En répétant cette procédure pour les $k \times s = 3 \times 3 = 9$ configurations possibles pour X et Y, on obtient le tableau de contingence théorique 11.7.

▼ Tableau 11.7 Tableau de contingence théorique

Montant \ Média	Emailing	Courriers	Appels	Total
50-100 euros	156,66	156,66	156,66	470
100-200 euros	163,33	163,33	163,33	490
plus de 200 euros	180	180	180	540
Total	500	500	500	1 500

La comparaison de ces deux tableaux de contingence se fait sur la base de la statistique du khi-deux. La réalisation de la statistique est égale à :

$$D_n(x,y) = \sum_{i=1}^{k} \sum_{j=1}^{s} \frac{(n_{i,j} - \widehat{N}_{i,j})^2}{\widehat{N}_{i,j}} \qquad (11.163)$$

$$= \frac{(220 - 156{,}66)^2}{156{,}66} + \ldots. + \frac{(350 - 180)^2}{180} \qquad (11.164)$$

$$= 447{,}42 \qquad (11.165)$$

La région critique du test pour un niveau de risque $\alpha = 10\ \%$ est définie par :

$$W = \left\{x,y : D_n(x,y) > G_4^{-1}(0{,}90)\right\} \qquad (11.166)$$

où $G_4(.)$ est la fonction de répartition de la loi du khi-deux à $(3-1) \times (3-1) = 4$ degrés de liberté. Sachant que $G_4^{-1}(0{,}90) = 7{,}77$, il vient :

$$W = \{x,y : D_n(x,y) > 7{,}77\} \qquad (11.167)$$

La réalisation de la statistique du khi-deux, égale à 447,42, appartient à la région critique. Par conséquent, on rejette l'hypothèse nulle d'indépendance entre le montant moyen des achats et le média pour un seuil de risque de 10 %. Cette conclusion confirme l'intuition que l'on pouvait avoir après avoir comparé les deux tableaux de contingence qui sont fort différents, surtout pour ce qui concerne les clients contactés par appel téléphonique.

Un des problèmes du test d'indépendance du khi-deux est que la statistique de test tend à être dégénérée lorsque *l'effectif théorique d'une classe est nul ou très faible*. Imaginons par exemple que pour la modalité a_1 de la variable X et la modalité b_1 de la variable Y, l'effectif théorique soit nul, *i.e.* $N_{1,1} = 0$. Dans ce cas, la statistique de test du khi-deux tend vers l'infini, puisque l'on divise l'un des éléments de la somme, *i.e.* $(n_{1,1} - N_{1,1})^2/N_{1,1}$ par zéro. Le même problème se pose lorsque l'effectif théorique est non nul, mais très faible. Le fait de diviser $(n_{i,j} - N_{i,j})^2$ par un effectif théorique $N_{i,j}$ proche de zéro induit que la réalisation de la statistique de test est très élevée. On a alors tendance à rejeter l'hypothèse nulle d'indépendance, et cela juste en raison de la faiblesse des effectifs théoriques de cette classe.

Remarque : Plusieurs auteurs ont proposé différents critères pour savoir si le test d'indépendance du khi-deux est valide. Une règle simple consiste à vérifier que pour toutes les classes les effectifs théoriques sont supérieurs à 5, c'est-à-dire que pour tout $i = 1,...,k$ et $j = 1,...,s$, on a $\widehat{N}_{i,j} > 5$. Dans le cas contraire, il convient de regrouper les classes sur X et/ou sur Y, de sorte à vérifier cette condition avant d'appliquer le test d'indépendance.

4.2 Tests d'adéquation du khi-deux

Définition 11.26

Le **test d'adéquation (ou d'ajustement) du khi-deux** permet de tester si des observations $(x_1,...,x_n)$ d'une variable aléatoire X sont issues d'une distribution que l'on spécifie. Les hypothèses nulle et alternative de ce test sont respectivement définies par :

$$\text{H}_0 : X \sim \text{D}(\theta) \quad \text{contre} \quad \text{H}_1 : X \text{ ne suit pas la loi D}(\theta) \tag{11.168}$$

où $\text{D}(\theta)$ désigne une distribution paramétrique de paramètre θ.

Par exemple, l'hypothèse nulle d'un test d'adéquation peut être $\text{H}_0 : X \sim \mathcal{P}(\lambda)$, où $\mathcal{P}(\lambda)$ désigne une loi de Poisson de paramètre λ, ou bien $\text{H}_0 : X \sim \mathcal{N}\left(m,\sigma^2\right)$. Le test d'adéquation du khi-deux peut en effet s'appliquer à des lois discrètes ou à des lois continues[3].

Le test d'adéquation est basé sur un **tableau de répartition empirique** des effectifs de l'échantillon $(x_1,...,x_n)$. On suppose que la variable X admet *s modalités distinctes*, notées $a_1,...,a_s$. Soit n_i, pour $i = 1,...,s$, le nombre d'individus dans l'échantillon pour lesquels on observe $X = a_i$, avec par définition $\sum_{i=1}^{s} n_i = n$. On obtient ainsi un tableau de répartition empirique similaire au tableau 11.8.

▼ Tableau 11.8 Tableau de répartition empirique

Variable X	$X = a_1$	..	$X = a_i$	..	$X = a_s$	Total
Effectifs empiriques	n_1	..	n_i	..	n_s	n

L'idée du test d'ajustement du khi-deux est de comparer ce tableau de répartition empirique (basé sur les observations de l'échantillon) à un tableau de répartition théorique que l'on obtiendrait si la variable X avait effectivement une distribution D. Si les deux tableaux sont quasiment identiques, alors on ne peut pas rejeter l'hypothèse nulle H_0 d'adéquation de la loi de X à D. En revanche, si les deux tableaux diffèrent, on rejette l'hypothèse nulle H_0 d'adéquation.

Comment construire le tableau de contingence théorique sous l'hypothèse nulle d'adéquation ? Pour cela, il faut connaître le nombre théorique, noté N_i, d'individus pour lesquels on devrait observer $X = a_i$ si la variable X avait effectivement une distribution D. Par définition, cet effectif théorique est égal à :

$$N_i = n \times \Pr(X = a_i) \tag{11.169}$$

3 Dans ce dernier cas, on découpe l'ensemble des valeurs que peut prendre X en s classes.

où $\Pr(X = a_i)$ désigne la probabilité théorique associée à la loi $D(\theta)$. A partir des effectifs N_i, pour $i = 1,\ldots,s$, on obtient un **tableau de répartition théorique** comme reporté dans le tableau 11.9.

▼ Tableau 11.9 **Tableau de répartition théorique**

Variable X	$X = a_1$	..	$X = a_i$	..	$X = a_s$	Total
Effectifs théoriques	N_1	..	N_i	..	N_s	n

Remarque : Par construction, la somme des effectifs théoriques sur toutes les modalités correspond à la taille d'échantillon, *i.e.* $\sum_{i=1}^{s} N_i = n$.

La comparaison des tableaux de répartition empirique et théorique se fait sur la base d'une statistique de test du khi-deux.

Définition 11.27

La **statistique de test d'adéquation du khi-deux**, notée C_n, est définie par :

$$C_n = \sum_{i=1}^{s} \frac{(n_i - N_i)^2}{N_i} \qquad (11.170)$$

Si les paramètres θ de la loi $D(\theta)$ sont **connus**, la statistique C_n admet une distribution exacte du khi-deux à $s - 1$ degrés de liberté sous l'hypothèse nulle :

$$C_n \underset{H_0}{\sim} \chi^2(s-1) \qquad (11.171)$$

La région critique du test est alors la suivante.

Définition 11.28

Si les paramètres θ de la loi $D(\theta)$ sont connus, la **région critique** du test d'adéquation du khi-deux pour un niveau de risque α est définie par :

$$W = \left\{x : C_n(x) > G_{s-1}^{-1}(1-\alpha)\right\} \qquad (11.172)$$

où $C_n(x)$ désigne la réalisation de la statistique C_n et $G_{s-1}(.)$ est la fonction de répartition de la loi du khi-deux à $s - 1$ degrés de liberté.

Ainsi, si la réalisation de la statistique du khi-deux est supérieure au fractile à $100 \times (1-\alpha)$ % de la loi du khi-deux à $s - 1$ degrés de liberté, on rejette l'hypothèse nulle $H_0 : X \sim D(\theta)$ pour un seuil de risque α.

Avant d'appliquer le test d'adéquation du khi-deux, il convient de s'assurer qu'aucune des modalités n'est associée à un effectif théorique nul ou trop faible. Une règle simple consiste à vérifier que $N_i > 5$ pour $i = 1,\ldots,s$. Dans le cas contraire, on regroupe certaines classes de sorte à vérifier cette condition.

Exemple

Afin d'ajuster au mieux la gestion du personnel des gares de péage, une société d'autoroute souhaite modéliser le nombre de voitures, noté X, se présentant à un péage par tranche de 30 minutes. On souhaite tester si la variable aléatoire X admet une distribution de Poisson de paramètre $\lambda = 2$, pour un niveau de risque de 5 %. On dispose pour cela d'un échantillon de 100 relevés consécutifs durant lesquels on a compté le nombre de voitures passant le péage. La répartition des effectifs empiriques est donnée dans le tableau 11.10.

▼ Tableau 11.10 Tableau de répartition empirique

Nombre de voitures (X)	0	1	2	3	4	5	6	7 et +	Total
Nombre de relevés (n_i)	15	25	25	12	10	8	4	1	100

Ce tableau se lit de la façon suivante : lors de 15 relevés, aucune voiture ne s'est présentée au péage, lors de 25 relevés, une seule voiture a été décomptée, etc. Afin de tester l'hypothèse nulle $H_0 : X \sim \mathcal{P}(2)$, construisons tout d'abord le tableau de répartition des effectifs théoriques. On sait que si X suit une loi $\mathcal{P}(\lambda)$, alors :

$$\Pr(X = x) = \exp(-\lambda)\frac{\lambda^x}{x!} \quad \forall x \in \mathbb{N} \tag{11.173}$$

Par conséquent sous H_0, l'effectif théorique associé à la modalité $X = 0$ est égal à :

$$N_0 = n \times \Pr(X = 0) = 100 \times \exp(-2) \times \frac{2^0}{0!} = 13{,}53 \tag{11.174}$$

De la même façon, l'effectif théorique associé à la modalité $X = 1$ est égal à :

$$N_1 = n \times \Pr(X = 1) = 100 \times \exp(-2) \times \frac{2^1}{1!} = 27{,}06 \tag{11.175}$$

Pour la modalité « 7 voitures ou plus », la probabilité $\Pr(X \geq 7)$ est définie par $1 - \sum_{x=0}^{6} \Pr(X = x)$. Par conséquent, l'effectif associé à cette classe est égal à :

$$N_{7+} = n - \sum_{i=0}^{6} N_6 = 100 - 13{,}53 - 27{,}06 - \ldots - 1{,}20 = 0{,}49 \tag{11.176}$$

On obtient ainsi tableau 11.11 donnant la répartition des effectifs théoriques sous H_0.

▼ Tableau 11.11 Tableau de répartition théorique

X	0	1	2	3	4	5	6	7 et +	Total
N_i	13,53	27,06	27,06	18,04	9,02	3,60	1,20	0,49	100

On constate que les modalités « 7 voitures ou plus » et « 4 voitures » ont des effectifs théoriques trop faibles, inférieurs à 5. Après regroupement des modalités 5, 6 et 7 et +, il reste $s = 6$ modalités (► tableau 11.12).

▼ Tableau 11.12 Effectifs empiriques et théoriques après regroupement

Nombre de voitures (X)	0	1	2	3	4	5 et +	Total
Nombre de relevés (n_i)	15	25	25	12	10	13	100
Effectifs théoriques (N_i)	13,53	27,06	27,06	18,04	9,02	5,29	100

La réalisation de la statistique d'ajustement du khi-deux est égale à :

$$C_n(x) = \sum_{i=1}^{6} \frac{(n_i - N_i)^2}{N_i} = \frac{(15 - 13{,}53)^2}{13{,}53} + \ldots + \frac{(13 - 5{,}29)^2}{5{,}29} = 13{,}83 \quad (11.177)$$

Sous H_0, la statistique C_n admet une distribution du khi-deux à $6 - 1 = 5$ degrés de liberté, *i.e.* le nombre de classes après regroupement moins 1. La région critique du test d'ajustement du khi-deux pour un niveau $\alpha = 5$ % est égale à :

$$W = \left\{x : C_n(x) > G_5^{-1}(0{,}95)\right\} \quad (11.178)$$

où $G_5(x)$ désigne la fonction de répartition de la loi du khi-deux à 5 degrés de liberté. On obtient ainsi :

$$W = \{x : C_n(x) > 11{,}07\} \quad (11.179)$$

La réalisation de la statistique de test, égale à 13,83, appartient à la région critique. Par conséquent, on rejette l'hypothèse nulle selon laquelle la variable X suit une loi de Poisson de paramètre 2, pour un seuil de significativité de 5 %.

Dans le cas où les paramètres θ sont *inconnus*, il convient de les estimer. Si l'on suppose que la loi D dépend de k paramètres, tels que $\theta = (\theta_1,\ldots,\theta_k)^\top$, la région critique du test devient :

Définition 11.29

Si les k paramètres θ de la loi $D(\theta)$ sont **estimés**, la statistique de test C_n admet une distribution exacte du khi-deux à $s - k - 1$ degrés de liberté sous l'hypothèse nulle :

$$C_n \underset{H_0}{\sim} \chi^2(s - k - 1) \quad (11.180)$$

où s désigne le nombre de modalités de X après regroupement. La **région critique** du test d'adéquation du khi-deux pour un niveau de risque α est alors définie par :

$$W = \left\{x : C_n(x) > G_{s-k-1}^{-1}(1 - \alpha)\right\} \quad (11.181)$$

où $C_n(x)$ désigne la réalisation de la statistique C_n et $G_{s-k-1}(.)$ est la fonction de répartition de la loi du khi-deux à $s - k - 1$ degrés de liberté.



2 questions à

Yoann **Grondin**

Analyste Statistique, EDF Commerce



Quel est votre parcours professionnel et votre mission actuelle chez EDF ?

À l'issue de mes études universitaires et de mon stage au sein de la direction marketing d'AXA, j'ai débuté ma carrière professionnelle en 2008 dans le département marketing de Bouygues Télécom. En 2011, j'ai été recruté comme analyste confirmé au sein du pôle « Analyse connaissance client » de la direction des services informatiques d'EDF Commerce. Je suis en charge de répondre aux problématiques métier de la direction EDF Commerce, principalement sur deux sujets que sont la digitalisation de la relation client (analyse des parcours multicanaux, score d'appétence au canal Internet...) et les départs à la concurrence (modèles de prévision, de durée de vie...) en apportant mon expertise statistique.

Quels sont les tests statistiques que vous utilisez dans votre activité au sein d'EDF ? Pouvez-vous nous expliquer leur utilité ?

L'utilisation des tests statistiques se pratique dans deux cadres d'analyse distincts. D'une part, lors de réalisation de scores ou d'autres méthodes de classification, les tests de liaison entre variables sont indispensables pour détecter rapidement les dépendances. On utilise ainsi les coefficients de corrélation de Pearson et des rangs de Spearman selon qu'il s'agisse de variables continues, discrètes ou ordinales, le test du khi-deux pour les variables nominales ou encore les tests de variance. D'autre part, dans le cadre des campagnes marketing, que ce soit pour la définition des cibles en amont ou pour l'évaluation de l'efficacité de la campagne en aval, les tests de comparaison de moyennes et de proportions sont fréquemment utilisés. ■

Les points **clés**

- Un test statistique est une règle de décision relative à une hypothèse nulle, établie sur la base d'un échantillon et permettant de contrôler les risques associés à la décision.
- Le risque de première espèce est le risque de rejeter à tort l'hypothèse nulle.
- La puissance d'un test correspond à la probabilité de rejeter l'hypothèse nulle lorsqu'elle n'est pas valide dans la population.
- La région critique correspond à l'ensemble des échantillons pour lesquels la réalisation de la statistique de test conduit au rejet de l'hypothèse nulle.
- La conclusion d'un test est une décision de rejet ou de non-rejet de l'hypothèse nulle pour un certain seuil de significativité (ou niveau).
- La p-value associée à une réalisation d'une statistique de test est le plus petit niveau pour lequel on peut rejeter l'hypothèse nulle.

ÉVALUATION

► Corrigés sur www.dunod.com

QCM

Pour chacune des questions suivantes, indiquez si les affirmations sont vraies ou fausses (il peut y avoir plusieurs réponses vraies pour chaque question).

1 Règle de décision d'un test statistique

a. La conclusion d'un test peut être le rejet de l'hypothèse alternative.

b. La conclusion d'un test peut être l'acceptation de l'hypothèse alternative.

c. La conclusion d'un test peut être l'acceptation de l'hypothèse nulle.

d. La conclusion d'un test peut être le rejet ou le non-rejet de l'hypothèse nulle pour un certain niveau de significativité.

e. Le rejet ou le non-rejet de l'hypothèse nulle ne dépend pas du niveau du test.

2 Région critique d'un test

a. La région critique caractérise l'ensemble des valeurs de la statistique de test pour lesquelles on rejette l'hypothèse nulle pour un niveau de risque donné.

b. Une statistique de test est une variable aléatoire.

c. Dans le cas d'un test paramétrique, la statistique de test est un estimateur du paramètre testé.

d. La valeur critique d'un test est établie à partir de la distribution de la statistique de test sous l'hypothèse nulle.

e. Si la réalisation de la statistique de test appartient à la région critique, on conclut au rejet de l'hypothèse nulle pour un seuil de significativité donné.

3 Niveau et puissance d'un test

a. Le niveau d'un test correspond à la probabilité de rejeter l'hypothèse alternative lorsque l'hypothèse nulle est vraie.

b. Un test de niveau 5 % a une valeur critique plus élevée qu'un test de niveau 10 %.

c. Le niveau d'un test est fixé par l'utilisateur.

d. La puissance correspond à la probabilité de rejeter l'hypothèse nulle alors que l'hypothèse alternative est vraie dans la population.

e. La puissance d'un test non biaisé tend vers l'unité lorsque la taille d'échantillon tend vers l'infini.

4 Test paramétrique et lemme de Neyman-Pearson

a. Dans le cas d'un test d'une hypothèse simple contre une hypothèse simple, le lemme de Neyman-Pearson permet de déterminer la région critique du test UPP pour un niveau donné.

b. Si l'on considère un test unilatéral $H_0 : \theta = \theta_0$ contre $H_1 : \theta < \theta_0$, la région critique du test peut être de la forme $W = \{x : \widehat{\theta}(x) > c\}$ où c est une valeur critique et $\widehat{\theta}$ un estimateur du paramètre θ.

c. Si l'on considère un test bilatéral du type $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$, la région critique du test peut être de la forme $W = \{x : \widehat{\theta}(x) - \theta_0 > c\}$ où c est une valeur critique et $\widehat{\theta}$ un estimateur du paramètre θ.

d. Le lemme de Neyman-Pearson permet de déterminer la région critique d'un test UPP bilatéral.

e. La région critique d'un test bilatéral de niveau α correspond à l'intersection des régions critiques des tests unilatéraux de niveau $\alpha/2$.

Sujets d'examen

5 Tests paramétriques (Université d'Orléans, 2013)

Soit X une variable aléatoire positive distribuée selon une *loi de Rayleigh* de paramètre σ^2. Sa fonction de densité est définie par :

$$f_X\left(x;\sigma^2\right) = \frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right) \quad \forall x \in \mathbb{R}^+ \qquad (11.182)$$

On souhaite tester la valeur de σ^2 à partir d'un n-échantillon $(X_1,\ldots,X_n)$ de taille $n = 100$, de variables i.i.d. de même loi que X. On cherche ainsi à tester :

$$\mathrm{H}_0 : \sigma^2 = \sigma_0^2 = 2 \quad \text{contre} \quad \mathrm{H}_1 : \sigma^2 = \sigma_1^2 = 2{,}1 \tag{11.183}$$

1. Montrer que la région critique du test UPP de niveau α peut s'écrire sous la forme $W = \{x | T_n(x) > c\}$ où c désigne une valeur critique et $T_n(x)$ est la réalisation de la statistique de test T_n, définie par :

$$T_n = \frac{1}{2n}\sum_{i=1}^{n} X_i^2 \tag{11.184}$$

2. Déterminer la valeur critique c du test UPP pour un niveau $\alpha = 5\ \%$, sachant que pour un échantillon de grande taille, on admet que :

$$T_n \overset{asy}{\approx} \mathcal{N}\left(\sigma^2, \frac{\sigma^4}{n}\right) \tag{11.185}$$

3. Déterminer la puissance du test de niveau $\alpha = 5\ \%$.
4. Montrer que le test est convergent.
5. Pour un échantillon d'observations, on obtient $\sum_{i=1}^{n} x_i^2 = 2\ 070$. Que conclure pour un seuil de significativité de 5 % ?
6. Quelle est la région critique du test unilatéral $\mathrm{H}_0 : \sigma^2 = 2$ contre $\mathrm{H}_1 : \sigma^2 > 2$ de niveau $\alpha = 5\ \%$?
7. Quelle est la région critique du test bilatéral $\mathrm{H}_0 : \sigma^2 = 2$ contre $\mathrm{H}_1 : \sigma^2 \neq 2$ de niveau $\alpha = 5\ \%$?

6 Tests paramétriques (HEC Lausanne, 2014)

On considère un échantillon $(X_1,\ldots,X_n)$ de variables aléatoires continues i.i.d. de même loi que X, où X est définie sur le support $X(\Omega) = [0,c]$ et admet une fonction de densité égale à :

$$f_X(x;\theta) = \frac{1}{\theta c^{1/\theta}} x^{\frac{1-\theta}{\theta}} \quad \forall x \in X(\Omega) \tag{11.186}$$

On suppose que la borne c est connue et que le paramètre θ est un paramètre positif inconnu que l'on cherche à estimer.

1. Écrire la log-vraisemblance associée à l'échantillon $(x_1,\ldots,x_n)$.
2. Montrer que l'estimateur $\widehat{\theta}$ du maximum de vraisemblance est défini par :

$$\widehat{\theta} = \ln(c) - \frac{1}{n}\sum_{i=1}^{n} \ln(X_i) \tag{11.187}$$

3. On admet que $\mathbb{E}(\ln(X_i)) = \ln(c) - \theta$. Montrer que l'estimateur $\widehat{\theta}$ est convergent.
4. Déterminer la distribution asymptotique de l'estimateur $\widehat{\theta}$.
5. On considère le test :

$$\mathrm{H}_0 : \theta = \theta_0 \quad \text{contre} \quad \mathrm{H}_1 : \theta = \theta_1 \tag{11.188}$$

avec $\theta_1 < \theta_0$. Montrer que la région critique du test UPP de niveau α a une forme générale du type :

$$\mathrm{W} = \left\{x : \widehat{\theta}(x) < A\right\} \tag{11.189}$$

où A est une constante déterminée par le niveau α et $\widehat{\theta}(x)$ désigne la réalisation de l'estimateur $\widehat{\theta}$ pour l'échantillon $(x_1,\ldots,x_n)$.

6. On admet que taille d'échantillon n est suffisamment importante pour supposer que :

$$\widehat{\theta} \overset{asy}{\approx} \mathcal{N}\left(\theta, \frac{\theta^2}{n}\right) \tag{11.190}$$

où θ désigne la vraie valeur du paramètre. Déterminer la valeur critique A du test UPP de niveau α.

7. On considère le test unilatéral :

$$\mathrm{H}_0 : \theta = \theta_0 \quad \text{contre} \quad \mathrm{H}_1 : \theta < \theta_0 \tag{11.191}$$

Déterminer la région critique du test UPP de niveau α.

8. On considère le test bilatéral :

$$\mathrm{H}_0 : \theta = \theta_0 \quad \text{contre} \quad \mathrm{H}_1 : \theta \neq \theta_0 \tag{11.192}$$

Déterminer la région critique du test de niveau α.

9. Montrer que le test de la question 8 est convergent.

CORRIGÉS

Les corrigés détaillés des QCM et de l'ensemble des autres exercices sont disponibles sur www.dunod.com, sur la page de l'ouvrage.

Chapitre 1

1 a. Faux b. Vrai c. Faux d. Vrai e. Vrai
2 a. Faux b. Faux c. Faux d. Vrai e. Vrai
3 a. Vrai b. Vrai c. Faux d. Faux e. Vrai
4 a. Faux b. Faux c. Faux d. Faux e. Vrai
5 a. Faux b. Vrai c. Faux d. Faux e. Vrai

Chapitre 2

1 a. Faux b. Vrai c. Faux d. Faux e. Vrai
2 a. Faux b. Faux c. Vrai d. Vrai e. Vrai
3 a. Faux b. Vrai c. Faux d. Faux e. Vrai
4 a. Faux b. Faux c. Faux d. Faux e. Vrai
5 a. Vrai b. Vrai c. Faux d. Faux e. Vrai

Chapitre 3

1 a. Vrai b. Faux c. Faux d. Vrai e. Vrai
2 a. Faux b. Faux c. Faux d. Vrai e. Vrai
3 a. Faux b. Faux c. Vrai d. Faux e. Faux
4 a. Faux b. Faux c. Vrai d. Faux e. Faux
5 a. Faux b. Faux c. Vrai d. Faux e. Faux

Chapitre 4

1 a. Faux b. Faux c. Faux d. Vrai e. Faux
2 a. Faux b. Faux c. Vrai d. Faux e. Vrai
3 a. Faux b. Vrai c. Vrai d. Faux e. Vrai
4 a. Faux b. Faux c. Faux d. Faux e. Vrai
5 a. Faux b. Faux c. Faux d. Vrai e. Faux

Chapitre 5

1 a. Vrai b. Vrai c. Vrai d. Faux e. Faux
2 a. Vrai b. Vrai c. Faux d. Faux e. Vrai
3 a. Faux b. Vrai c. Vrai d. Vrai e. Faux
4 a. Faux b. Vrai. c. Vrai d. Vrai e. Faux

Chapitre 6

1 a. Faux b. Faux c. Vrai d. Vrai e. Faux
2 a. Vrai b. Faux c. Faux d. Vrai e. Faux
3 a. Faux b. Vrai c. Faux d. Faux e. Vrai
4 a. Faux b. Faux c. Vrai d. Vrai e. Vrai

Chapitre 7

1 a. Faux b. Vrai c. Vrai d. Vrai e. Vrai
2 a. Vrai b. Faux c. Faux d. Faux e. Vrai
3 a. Vrai b. Faux c. Vrai d. Vrai e. Vrai
4 a. Faux b. Vrai c. Faux d. Vrai e. Faux

Chapitre 8

1 a. Faux b. Vrai c. Vrai d. Vrai e. Faux
2 a. Faux b. Vrai c. Faux d. Faux e. Vrai
3 a. Faux b. Vrai c. Vrai d. Faux e. Vrai
4 a. Vrai b. Vrai c. Vrai d. Vrai e. Faux

Chapitre 9

1 a. Faux b. Faux c. Vrai d. Faux e. Vrai
2 a. Vrai b. Vrai c. Faux d. Faux e. Faux
3 a. Vrai b. Faux c. Vrai d. Vrai e. Faux
4 a. Vrai b. Vrai c. Vrai d. Vrai e. Faux

Chapitre 10

1 a. Vrai b. Faux c. Faux d. Faux e. Vrai
2 a. Faux b. Vrai c. Faux d. Vrai e. Faux
3 a. Vrai b. Vrai c. Vrai d. Vrai e. Faux
4 a. Faux. b. Faux c. Vrai d. Faux e. Faux

Chapitre 11

1 a. Faux b. Faux c. Faux d. Vrai e. Faux
2 a. Vrai b. Vrai c. Faux d. Vrai e. Vrai
3 a. Faux b. Faux d. Vrai e. Faux
4 a. Vrai b. Faux c. Faux d. Faux e. Faux

Bibliographie

AMEMIYA T., *Advanced Econometrics*, Harvard University Press, 1985.

BERTHIER, J.-P., « Introduction à la pratique des indices statistiques », Document de travail n°M0503 de la Direction des statistiques démographiques et sociales, INSEE, 2005.

CHAUVET-PEYRARD A., « Indices des prix à la consommation 1998-2012 selon la région d'habitation des ménages et selon la taille de la commune de résidence », Document de travail n°F1306 de la Direction des statistiques démographiques et sociales, INSEE, 2013.

GOURIÉROUX, C. et MONFORT, A., *Statistique et modèles économétriques*, Economica, 1989.

GOURIÉROUX, C. et MONFORT, A., *Séries temporelles et modèles dynamiques*, Economica, 1995.

GRAIS, B., *Statistique descriptive*, Dunod, 2003.

GREENE, W., *Econométrie*, Pearson Education, 2005.

GUÉDÈS, D., « Impact des ajustements de qualité dans le calcul de l'indice des prix à la consommation », Document de travail n°F0404 de la Direction des statistiques démographiques et sociales, INSEE, 2004.

MIGNON, V., *Econométrie. Théorie et applications*, Economica, 2008.

PY, B., *Statistique descriptive : nouvelle méthode pour bien comprendre et réussir*, Economica, 2007.

WINTERS, P.R., « Forecasting sales by exponentially weighted moving average », *Management Science*, pp. 324–342, 1960.

Index



M

N

P

Q

R

S

T

U

V

Z

CORRIGÉS

Chapitre 1

1 Variables discrètes et variables continues

a. Faux, une variable discrète peut aussi prendre des valeurs négatives.
b. Vrai.
c. Faux, le chiffre d'affaires est une variable continue.
d. Vrai.
e. Vrai.

2 Caractères et modalités

a. Faux. Par exemple, le caractère « sexe » n'a que deux modalités.
b. Faux, les modalités doivent être incompatibles.
c. Faux, car si la taille est bien un caractère quantitatif, l'état matrimonial est un caractère qualitatif.
d. Vrai.
e. Vrai.

3 Graphiques et centre de classe

a. Vrai.
b. Vrai.
c. Faux, il s'agit de la courbe des fréquences cumulées.
d. Faux, il est nécessaire de corriger les amplitudes afin que l'aire de chaque rectangle composant l'histogramme soit bien proportionnelle à l'effectif (ou la fréquence).
e. Vrai.

4 Le mode

a. Faux, il s'agit de la médiane.
b. Faux.
c. Faux, il s'agit de la moyenne.
d. Faux, le mode est une caractéristique de tendance centrale.
e. Vrai.

5 Étude de la répartition des notes : 18, 15, 8, 12, 8, 15, 4

a. Faux, il convient de pondérer par les effectifs : $\frac{18 + 15 \times 2 + 8 \times 2 + 12 + 4}{7} = 11{,}42$.
b. Vrai.
c. Faux, la distribution est bimodale, les deux valeurs du mode étant 8 et 15.
d. Faux, l'étendue est égale à 18 – 4 = 14.
e. Vrai, le moment simple d'ordre 1 étant égal à la moyenne.

6 Caractéristiques d'une distribution

Le tableau 1.1 reporte les différents calculs nécessaires pour répondre aux diverses questions posées.

Tableau 1.1 Location de voitures

Classes	n_i	x_i	f_i	F_i	$n_i x_i$	$n_i x_i^2$
[0,10[	2	5	0,01	0,01	10	50
[10,20[	19	15	0,11	0,13	285	4 275
[20,30[	28	25	0,17	0,30	700	17 500
[30,40[	54	35	0,33	0,62	1 890	66 150
[40,50[	31	45	0,19	0,81	1 395	62 775
[50,60[	21	55	0,13	0,93	1 155	63 525
[60,70[	11	65	0,07	1,00	715	46 475
Total	**166**		**1**		**6 150**	**260 750**

1. La variable étant groupée en classes, il s'agit d'une variable continue.
2. Les résultats sont reportés dans le tableau 1.1, x_i désignant les centres de classes, f_i les fréquences et F_i les fréquences cumulées.
3. Il est possible de représenter la série au moyen d'un histogramme.
4. La classe modale est celle pour laquelle la fréquence est la plus élevée, il s'agit donc de la classe [30,40[. On peut calculer la valeur du mode au moyen de la relation (1.14) :

$$Mode = 30 + 10 \times \frac{0{,}33 - 0{,}17}{(0{,}33 - 0{,}17) + (0{,}33 - 0{,}19)} = 35{,}42$$

La valeur du mode est donc égale à 35 voitures.
5. La moyenne peut être calculée à l'aide de l'équation (1.24) :

$$\bar{x} = \frac{1}{166} \times 6\,150 = 37{,}05$$

En moyenne, 37 voitures sont louées par jour.
6. On utilise la colonne des fréquences cumulées afin de repérer la classe médiane, il s'agit ainsi de la classe [30,40[. La valeur précise de la médiane est obtenue à l'aide de l'équation (1.17) :

$$M = 30 + \frac{10}{0{,}33} \times [0{,}5 - 0{,}30]$$

La valeur médiane est ainsi égale à 36,30 : il y a autant de sociétés louant moins de 36 voitures que de sociétés en louant plus.
7. On calcule l'écart-type corrigé :

$$S_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} n_i x_i^2 - \frac{N}{N-1}\bar{x}^2}$$

soit :

$$S_x = \sqrt{\frac{1}{166-1} \times 260\,750 - \frac{166}{166-1} \times 37{,}05^2}$$

L'écart-type est ainsi égal à 14,12 voitures.

7 Sujet d'examen

▼ **Tableau 1.2** Répartition des salariés par tranche de salaire

Classes de salaires	n_i	x_i	a_i	f_i	F_i	n_ix_i	$\frac{n_ix_i}{\sum_{i=1}^{5} n_ix_i}$	$\sum_{i=1}^{5} \frac{n_ix_i}{\sum_{i=1}^{5} n_ix_i}$
[1200,1400[	3	1 300	200	0,015	0,015	3 900	0,011	0,011
[1400,1600[	6	1 500	200	0,03	0,045	9 000	0,027	0,038
[1600,1800[	182	1 700	200	0,91	0,955	309 400	0,909	0,947
[1800,2000[	5	1 900	200	0,025	0,98	9 500	0,028	0,975
[2000,2200[	4	2 100	200	0,02	1	8 400	0,025	1
Total	**200**			**1**		**340 200**	**1**	

1. L'effectif total de l'entreprise est donné par :

$$N = \sum_{i=1}^{5} n_i = 3 + 6 + 182 + 5 + 4 = 200$$

2. Les valeurs demandées sont reportées dans le tableau 1.2.

3. La moyenne est donnée par :

$$\bar{x} = \frac{1}{200} \times 340\,200 = 1701$$

Le salaire moyen au sein de l'entreprise considérée est de 1701 euros.

4. La classe modale est la classe pour laquelle l'effectif (ou la fréquence) est le plus élevé, il s'agit en conséquence de la classe [1600,1800[. La valeur du mode est obtenue par la formule :

$$Mode = e_{i-1} + a_m \times \frac{d_1}{d_1 + d_2}$$

où e_{i-1} désigne la valeur de l'extrémité inférieure de la classe modale (1600), a_m l'amplitude de cette même classe (200), d_1 la différence entre l'effectif de la classe modale (182) et l'effectif de la classe précédente (6) et d_2 la différence entre l'effectif de la classe modale (182) et l'effectif de la classe suivante (5). Dans le cas de notre entreprise, le mode est ainsi égal à 1 699,72 euros :

$$Mode = 1\,600 + 200 \times \frac{182 - 6}{(182 - 6) + (182 - 5)} = 1\,699{,}72$$

5. L'étendue est donnée par la différence entre la valeur maximale et la valeur minimale prise par le salaire. Dans notre cas, les observations étant groupées par classes, l'étendue est égale à la différence entre l'extrémité supérieure de la dernière classe et l'extrémité inférieure de la première classe, soit :

$$Etendue = 2\,200 - 1\,200 = 1\,000$$

L'étendue des salaires au sein de l'entreprise considérée est ainsi égale à 1 000 euros.

6. Au regard des valeurs prises par les fréquences cumulées, on constate que la proportion de salariés gagnant moins de 1 600 euros est de 4,5 % et que celle touchant moins de 1 800 euros est de 95,5 %. On en déduit que la valeur de la médiane est comprise entre 1 600 et 1 800 euros et que la classe médiane est la classe [1600,1800[. La valeur de la médiane est donnée par :

$$M = e_{i-1} + \frac{a_i}{f_i} \times [0{,}5 - F_{i-1}]$$

où e_{i-1} est l'extrémité inférieure de la classe médiane (1600), a_i l'amplitude de la classe médiane (200), f_i la fréquence de la classe médiane (0,91) et F_{i-1} désigne la fréquence cumulée de la classe au dessus de la classe médiane dans le tableau (0,045). D'où :

$$M = 1\,600 + \frac{200}{0{,}91} \times [0{,}5 - 0{,}045] = 1\,700$$

La médiane est ainsi égale à 1 700 euros : dans l'entreprise considérée, il y a autant de salariés gagnant moins de 1 700 euros par mois que de salariés percevant un salaire mensuel supérieur à 1 700 euros.

7. Trois caractéristiques de tendance centrale ont été calculées : la moyenne égale à 1 701 euros, le mode égal à 1 699,72 euros et la médiane valant 1 700 euros. Ces trois valeurs étant très proches, on en déduit que la distribution des salaires au sein de l'entreprise est symétrique.

8. La détermination de la classe médiale est similaire à celle de la classe médiane, le calcul étant basé non plus sur les seuls effectifs n_i mais sur le produit n_ix_i représentant la masse salariale. Par un raisonnement identique et au regard des valeurs obtenues dans la dernière colonne du tableau 1.2, on en déduit que la classe médiale est la classe [1600,1800[. De même, la valeur de la médiale, notée Ml et exprimée en euros, est donnée par :

$$Ml = 1\,600 + \frac{200}{0{,}909} \times [0{,}5 - 0{,}038] = 1\,701{,}65$$

9. On constate que les valeurs de la médiane et de la médiale sont très proches. Plus précisément, on a :

$$\frac{Ml - M}{Etendue} = \frac{1\,701{,}65 - 1\,700}{1\,000} = 0{,}00\,165$$

Le rapport calculé est ainsi très proche de zéro, l'écart entre la médiale et la médiane étant très faible par rapport à l'étendue. Cela correspond à une concentration nulle des salaires, c'est-à-dire à une parfaite équipartition. Ce résultat était prévisible au regard des données figurant dans le tableau 1.13 : la distribution des observations fait en effet ressortir que la moyenne est très proche de 1 700 euros, témoignant d'une très faible dispersion des salaires autour du salaire moyen.

Chapitre 2

1 Distributions marginales et conditionnelles

a. Faux, la distribution marginale d'une variable ne tient compte que de l'information contenue dans cette variable.

b. Vrai.

c. Faux, la somme des fréquences, quel que soit le type de fréquence considéré, est toujours égale à 1.

d. Faux, il existe toujours une relation entre fréquences marginales et conditionnelles (▸ équation (2.15)).

e. Vrai.

2 Liaison entre deux variables

a. Faux, il s'agit d'un cas de corrélation négative.

b. Faux, deux variables corrélées négativement évoluent en sens contraire.
c. Vrai.
d. Vrai.
e. Vrai.

3 Droite de régression et ajustement

a. Faux.
b. Vrai.
c. Faux, il s'agit de minimiser la somme des carrés des écarts et non la somme des écarts.
d. Faux, b est l'ordonnée à l'origine et a est la pente de la droite de régression.
e. Vrai.

4 Coefficient de corrélation linéaire

a. Faux, il s'agit d'une corrélation négative.
b. Faux, un coefficient de corrélation est compris entre −1 et 1.
c. Faux, il ne faut pas confondre corrélation et causalité.
d. Faux, les variables peuvent être corrélées de façon non-linéaire.
e. Vrai.

5 Analyse de la variance et coefficient de détermination

a. Vrai.
b. Vrai.
c. Faux, c'est le rapport entre la variance expliquée et la variance totale.
d. Faux, un coefficient de détermination est compris entre 0 et 1.
e. Vrai.

6 Étude de la liaison taux de chômage/taux d'inflation

1. La figure 2.10 montre que le nuage de points est relativement allongé et semble se répartir autour d'une droite d'allure décroissante. Cela laisse présager l'existence d'une liaison négative entre les deux variables. Un tel résultat est cohérent d'un point de vue économique en vertu de la relation de Phillips selon laquelle les deux variables tendent à évoluer en sens contraire.

En toute rigueur, notons que nous considérons ici une relation linéaire entre le taux de chômage et le taux d'inflation alors que la courbe de Phillips renvoie à une relation non-linéaire entre les deux variables.

2. Rappelons que la covariance entre deux variables x et y est donnée par :

$$Cov(x,y) = \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \bar{x}\bar{y}$$

D'où :

$$Cov(x,y) = \frac{1}{30} \times 881{,}01 - 6{,}71 \times 4{,}74 = -2{,}43$$

La covariance est négative, comme attendu.

3. La valeur du coefficient de corrélation linéaire entre les deux variables est donnée par :

$$r(x,y) = \frac{Cov(x,y)}{\sigma_x \sigma_y}$$

D'où :

$$r(x,y) = \frac{-2{,}43}{\sqrt{4{,}64} \times \sqrt{1{,}51}} = -0{,}92$$

Le coefficient de corrélation linéaire entre le taux de chômage et le taux d'inflation est négatif et proche de −1. Il existe donc bien une relation décroissante entre les deux variables, conformément à la relation de Phillips.

4. Les valeurs des coefficients de la droite de régression sont donnés par :

$$a = \frac{Cov(x,y)}{V(x)} = \frac{-2{,}43}{4{,}64} = -0{,}52$$

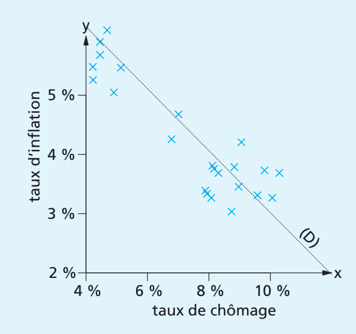
et, sachant que la droite de régression passe par le point moyen :

$$b = \bar{y} - a\bar{x} = 4{,}74 + 0{,}52 \times 6{,}71 = 8{,}23$$

On en déduit donc que la droite de régression a pour expression :

$$\hat{y} = -0{,}52x + 8{,}23$$

Le coefficient de pente de la droite de régression est bien négatif, confirmant à nouveau la relation inverse entre le taux d'inflation et le chômage. La figure 2.1 reproduit l'ajustement du nuage de points par cette droite de régression.



▲ **Figure 2.1 Ajustement du nuage de points par la droite de régression.**

5. On peut calculer le coefficient de détermination comme suit :

$$R^2 = \frac{Cov(x,y)^2}{V(x)V(y)} = \frac{(-2{,}43)^2}{4{,}64 \times 1{,}51} = 0{,}84$$

On en déduit que le pourcentage de variance expliquée par la régression est de 84 %.

7 Sujet d'examen

1. Les coefficients de variation CV, exprimés en pourcentage, sont donnés par le rapport entre l'écart-type et la moyenne pour chacune des trois séries. On a donc : $CV_{EUR} = 104{,}1565$, $CV_{DKK} = 6{,}5569$ et $CV_{ATS} = 5{,}2164$. La série la plus volatile est donc la série EUR, suivie par DKK puis ATS.

2. On rappelle que les coefficients de corrélation sont donnés par :

$$r(EUR,DKK) = \frac{Cov(EUR,DKK)}{\sigma_{EUR}\sigma_{DKK}}$$

$$r(ATS,DKK) = \frac{Cov(ATS,DKK)}{\sigma_{ATS}\sigma_{DKK}}$$

On obtient ainsi :

$$r(EUR,DKK) = \frac{\frac{1}{172}(-37{,}3106) - (-0{,}1227) \times 1{,}8942}{0{,}1278 \times 0{,}1242} = 0{,}9763$$

$$r(ATS,DKK) = \frac{\frac{1}{172}(817{,}2116) - 2{,}4998 \times 1{,}8942}{0{,}1304 \times 0{,}1242} = 0{,}9947$$

On constate que les deux coefficients de corrélation sont positifs : les deux variables considérées évoluent dans le même sens. Ils sont en outre très proches de 1, ce qui témoigne d'une forte corrélation entre les deux variables ; la corrélation la plus forte étant obtenue entre les variables DKK et ATS.

3. **a.** L'application de la méthode des moindres carrés ordinaires conduit aux résultats suivants :

$$a = \frac{Cov(EUR,DKK)}{V(EUR)} = \frac{\frac{1}{172}(-37{,}3106) - (-0{,}1227) \times 1{,}8942}{(0{,}1278)^2} = 0{,}9488$$

$$b = 1{,}8942 - 0{,}9488 \times (-0{,}1227) = 2{,}0106$$

La droite de régression s'écrit donc :
$\widehat{DKK}_t = 0{,}9488 \times EUR_t + 2{,}0106$

b. Toutes choses égales par ailleurs, une augmentation de 1 % de EUR s'accompagne d'une hausse de 0,9488 % du taux de change du dollar vis-à-vis de la couronne danoise.

c. L'équation d'analyse de la variance est donnée par :

$$\text{Variance totale} = \text{Variance expliquée} + \text{Variance résiduelle} \quad (2.1)$$

La variance totale est égale à $0{,}1242^2$ et la variance résiduelle vaut $0{,}1169/172 = 0{,}0007$. On en déduit donc que la variance expliquée est égale à : $0{,}1242^2 - 0{,}0006 = 0{,}0147$. La part de la variance résiduelle dans la variance totale est en conséquence égale à $0{,}0007/0{,}1242^2 = 0{,}0454$, soit 4,54 %.

d. Le coefficient de détermination est donné par le rapport entre la variance expliquée et la variance totale, soit :

$$R^2 = \frac{0{,}0147}{(0{,}1242)^2} = 0{,}9530$$

La variable EUR explique donc 95,30 % des variations du taux de change du dollar vis-à-vis de la couronne danoise.

e. Si l'on multiplie par 10 toutes les valeurs observées des variables EUR et DKK, cela n'a aucun impact sur la pente de la droite de régression. En effet, soient : $DKK'_t = 10 \times DKK_t$ et $EUR'_t = 10 \times EUR_t$. On peut écrire :

$$\frac{DKK'_t}{10} = a \times \frac{EUR'_t}{10} + b$$

soit encore :

$$DKK'_t = a \times EUR'_t + 10b$$

Si l'on ajoute 10 à chacune des valeurs observées des deux variables, le coefficient de la pente de la droite de régression n'est pas modifié. Posons : $DKK''_t = 10 + DKK_t$ et $EUR''_t = 10 + EUR_t$. On a :

$$DKK''_t - 10 = a(EUR''_t - 10) + b$$

soit encore :

$$DKK''_t = a \times EUR''_t + 10 - 10a + b$$

4. On sait que la variance totale est égale à $0{,}1242^2$ et que la variance expliquée vaut 0,0153. On en déduit, d'après l'équation d'analyse de la variance, la valeur de la variance résiduelle : $0{,}1242^2 - 0{,}0153 = 1{,}2564.10^{-4}$. La somme des carrés des résidus est en conséquence égale à : $172 \times 1{,}2564.10^{-4} = 0{,}0216$. Le coefficient de détermination est quant à lui donné par :

$$R^2 = \frac{0{,}0153}{(0{,}1242)^2} = 0{,}9919$$

On en déduit que la variable ATS explique 99,19 % de la variation de DKK.

5. Les deux modèles ayant la même variable expliquée, DKK, il est possible de les comparer. Le coefficient de détermination du deuxième modèle est supérieur à celui du premier modèle. La variance résiduelle est plus faible dans le deuxième modèle. Au regard de ces résultats, on retiendra donc le deuxième modèle pour expliquer les fluctuations de DKK.

Chapitre 3

1 Propriétés sur les indices

a. Vrai.

b. Faux. On a en effet : $I^g_{t/t'} = \dfrac{I^g_{t/0}}{I^g_{t'/0}}$.

c. Faux, aucun des deux indices ne vérifie ces propriétés.

d. Vrai.

e. Vrai.

2 Indices synthétiques

a. Faux, il s'agit de l'indice des prix de Laspeyres.

b. Faux, l'indice de Paasche est une moyenne harmonique.

c. Faux, les indices élémentaires doivent être basés à la même date.

d. Vrai, la structure du panier est celle de la date courante.

e. Vrai, la structure du panier est celle de la date de référence.

3 Évolution du prix d'un quotidien de la presse écrite

a. Faux.
b. Faux.
c. Vrai. En effet, en utilisant les propriétés de circularité et de réversibilité, on a :

$$I_{2013/2005} = I_{2013/2000} \times I_{2000/2005} = \frac{I_{2013/2000}}{I_{2005/2000}} = \frac{1{,}30}{1{,}50} = 0{,}867$$

d. Faux.
e. Faux.

4 Évolution d'un indice de prix

a. Faux.
b. Faux.
c. Vrai.
d. Faux.
e. Faux.

5 Formule d'un indice synthétique

a. Faux.
b. Faux.
c. Vrai.
d. Faux.
e. Faux.

6 Indices élémentaires, synthétiques et effet qualité

▼ **Tableau 3.1** Consommation de pommes d'un ménage : calcul des indices

Variété i	p_0^i	q_0^i	p_t^i	q_t^i	$p_0^i q_0^i$	$p_t^i q_t^i$	$p_0^i q_t^i$	$p_t^i q_0^i$
Golden	2,8	10	3,1	15	28	46,5	42	31
Pink Lady	3,9	5	4,4	15	19,5	66	58,5	22
Royal Gala	2,1	15	2,2	10	31,5	22	21	33
Somme		30		40	79	134,5	121,5	86

▼ **Tableau 3.2** Consommation de pommes d'un ménage : coefficients de pondération

Variété i	α_0^i	α_t^i	$\alpha_t^i \frac{p_0^i}{p_t^i}$	$\alpha_0^i \frac{p_t^i}{p_0^i}$	$\alpha_0^i \frac{q_t^i}{q_0^i}$	$\alpha_t^i \frac{q_0^i}{q_t^i}$
Golden	0,35	0,35	0,31	0,39	0,53	0,23
Pink Lady	0,25	0,49	0,43	0,28	0,74	0,16
Royal Gala	0,40	0,16	0,16	0,42	0,27	0,25
Somme	1	1	0,90	1,09	1,54	0,64

1. Le prix moyen du kilogramme de pommes en 2010 et en 2013 est donné par :

$$\bar{p}_0 = \frac{\Sigma_i p_0^i q_0^i}{\Sigma_i q_0^i} = \frac{79}{30} = 2{,}63$$

$$\bar{p}_t = \frac{\Sigma_i p_t^i q_t^i}{\Sigma_i q_t^i} = \frac{134{,}5}{40} = 3{,}36$$

On en déduit l'indice élémentaire du prix moyen :

$$I_{t/0}^g(\bar{p}) = \frac{\bar{p}_t}{\bar{p}_0} = \frac{3{,}36}{2{,}63} = 1{,}278$$

Le prix moyen du kilogramme de pommes a augmenté de 27,8 % entre 2010 et 2013.

2. L'indice élémentaire des quantités s'obtient comme suit :

$$I_{t/0}^g(q) = \frac{q_t}{q_0} = \frac{40}{30} = 1{,}333$$

et montre que la consommation de pommes de notre ménage s'est accrue de 33,3 % entre 2010 et 2013.

3. Le tableau 3.1 fournit les calculs nécessaires à l'obtention des indices des prix de Laspeyres et de Paasche donnés par les relations suivantes :

$$L_{t/0}^g(\bar{p}) = \frac{\Sigma_i p_t^i q_0^i}{\Sigma_i p_0^i q_0^i} = \frac{86}{79} = 1{,}089$$

$$P_{t/0}^g(\bar{p}) = \frac{\Sigma_i p_t^i q_t^i}{\Sigma_i p_0^i q_t^i} = \frac{134{,}5}{121{,}5} = 1{,}107$$

Il est également possible d'effectuer les calculs à l'aide des coefficients de pondération (► tableau 3.2) :

– Pour l'indice de Laspeyres, rappelons que les poids sont donnés par :

$$\alpha_0^i = \frac{p_0^i q_0^i}{\Sigma_i p_0^i q_0^i}$$

D'où :

$$L_{t/0}^g(\bar{p}) = \Sigma_i \alpha_0^i \frac{p_t^i}{p_0^i} = 1{,}089$$

– Pour l'indice de Paasche, rappelons que les poids sont donnés par :

$$\alpha_t^i = \frac{p_t^i q_t^i}{\Sigma_i p_t^i q_t^i}$$

et l'on a donc :

$$P_{t/0}^g(\bar{p}) = \frac{1}{\Sigma_i \alpha_t^i \frac{p_0^i}{p_t^i}} = \frac{1}{0{,}903} = 1{,}107$$

On retrouve bien entendu les mêmes valeurs. La hausse du prix moyen du kilogramme de pommes entre 2010 et 2013 est comprise entre 8,9 % et 10,7 % selon que l'on retient l'indice de Laspeyres ou de Paasche.

4. Les indices des quantités de Laspeyres et de Paasche s'obtiennent à l'aide des calculs figurant dans le tableau 3.1 :

$$L_{t/0}^g(q) = \frac{\Sigma_i p_0^i q_t^i}{\Sigma_i p_0^i q_0^i} = \frac{121{,}5}{79} = 1{,}538$$

$$P_{t/0}^g(q) = \frac{\Sigma_i p_t^i q_t^i}{\Sigma_i p_t^i q_0^i} = \frac{134{,}5}{86} = 1{,}564$$

5. Les indices de Fisher sont donnés par :

$$F_{t/0}^g(\bar{p}) = \sqrt{L_{t/0}^g(\bar{p}) P_{t/0}^g(\bar{p})} = \sqrt{1{,}089 \times 1{,}107} = 1{,}098$$

et

$$F_{t/0}^g(q) = \sqrt{L_{t/0}^g(q) P_{t/0}^g(q)} = \sqrt{1{,}538 \times 1{,}564} = 1{,}551$$

On retrouve bien le fait que la valeur de l'indice de Fisher est comprise entre les valeurs prises par les indices de Laspeyres et de Paasche. En revanche, on constate que la valeur prise par l'indice de Paasche est supérieure à celle de l'indice de Laspeyres, contrairement au cas fréquemment rencontré. Ce résultat n'est toutefois pas surprenant au regard des valeurs associées aux coefficients de pondération α_0^i et α_t^i. En effet, rappelons que l'inégalité selon laquelle l'indice de Paasche est généralement inférieur à l'indice de Laspeyres repose sur l'hypothèse de coefficients de pondération égaux entre les deux indices. Dans le cadre de notre exemple, on constate que les coefficients de pondération diffèrent fortement entre les deux dates, notamment pour les variétés Pink Lady et Royal Gala. On notera que les indices élémentaires des prix associés aux variétés Golden et surtout Pink Lady sont ceux qui ont le plus augmenté, ces deux variétés étant celles pour lesquelles les coefficients de pondération se sont accrus. Dans ces conditions, l'inégalité n'a plus de raison d'être vérifiée.

6. Calculons les indices de qualité :

$$S_t = \frac{I_{t/0}^g(\bar{p})}{L_{t/0}^g(\bar{p})} = \frac{P_{t/0}^g(q)}{I_{t/0}^g(q)} = \frac{1{,}278}{1{,}089} = \frac{1{,}564}{1{,}333} = 1{,}173$$

et :

$$S'_t = \frac{I_{t/0}^g(\bar{p})}{P_{t/0}^g(\bar{p})} = \frac{L_{t/0}^g(q)}{I_{t/0}^g(q)} = \frac{1{,}278}{1{,}107} = \frac{1{,}538}{1{,}333} = 1{,}153$$

L'indice de valeur est donné par :

$$I_{t/0}^V = L_{t/0}^g(\bar{p}) \times S_t \times I_{t/0}^g(q) = P_{t/0}^g(\bar{p}) \times S'_t \times I_{t/0}^g(q)$$

soit, avec nos données :

$$I_{t/0}^V = 1{,}089 \times 1{,}173 \times 1{,}333 = 1{,}107 \times 1{,}153 \times 1{,}333 = 1{,}703$$

La hausse de 70,3 % de la valeur du panier de fruits du ménage se décompose comme suit :

– une hausse du prix moyen comprise entre 8,9 % et 10,7 % ;

– une hausse de la quantité consommée de 33,3 % ;

– un effet de structure, pour 15,3 % à 17,3 %.

Cet effet de structure provient d'une baisse de la consommation de la variété la moins chère (Royal Gala) au profit de pommes plus onéreuses, la Golden et surtout la Pink Lady. Ce résultat était attendu au vu des valeurs prises par les coefficients de pondération : alors qu'en 2010 le coefficient de pondération le plus élevé était associé à la variété Royal Gala, il est associé à la variété Pink Lady en 2013.

7 Sujet d'examen

▼ Tableau 3.3 Prix et quantités des biens *A*, *B*, *C* et *D*

Bien h	p_0^h	p_t^h	q_0^h	q_t^h	$I_{t/0}^h$	$p_t^h q_0^h$	$p_0^h q_0^h$	$p_t^h q_t^h$	$p_0^h q_t^h$
A	169	610	210	160	3,61	128 100	35 490	97 600	27 040
B	81	265	220	230	3,27	58 300	17 820	60 950	18 630
C	1 023	2 470	30	18	2,41	74 100	30 690	44 460	18 414
D	32	64	470	470	2,00	30 080	15 040	30 080	15 040
Total						290 580	99 040	233 090	79 124

1. Les quantités consommées pour chacun des groupes de biens en 2014 (notées q_t^h) sont reportées dans la cinquième colonne du tableau 3.3. Elles s'obtiennent comme suit :

– Pour le bien A : $210 \times (1 - 0{,}238) = 160$

– Pour le bien B : $220 \times (1 + 0{,}0454) = 230$

– Pour le bien C : $30 \times (1 - 0{,}40) = 18$

– Pour le bien D : $470 \times (1 + 0) = 470$

2. Les indices élémentaires des prix pour chaque groupe de biens, base 1 en 2000, notés $I_{t/0}^h$, sont donnés par :

$$I_{t/0}^h = \frac{p_t^h}{p_0^h}$$

et les valeurs correspondantes sont reportées dans la sixième colonne du tableau 3.3. Les prix des 4 biens ont augmenté entre 2000 et 2014. Le bien dont le prix a le plus augmenté est le bien A, celui dont le prix a le moins augmenté est le bien D.

3. Les calculs nécessaires à la détermination des indices des prix de Laspeyres et de Paasche figurent dans les quatre dernières colonnes du tableau 3.3. On a ainsi :

$$L_{t/0}^h(p) = \frac{290\,580}{99\,040} = 2{,}934$$

$$P_{t/0}^h(p) = \frac{233\,090}{79\,124} = 2{,}946$$

On en déduit la valeur de l'indice des prix de Fisher :

$$F_{t/0}^h(p) = \sqrt{293{,}4 \times 294{,}6} = 2{,}940$$

On relève ainsi une augmentation des prix de l'ordre de 194 % entre 2000 et 2014. On constate que la valeur de l'indice de Fisher est bien comprise entre les valeurs des deux autres indices.

8 Sujet d'examen

▼ Tableau 3.4 Prix, quantités et indices élémentaires – Biens *a,b,c*

h	p_0^h	q_0^h	p_t^h	q_t^h	$p_0^h q_0^h$	$p_t^h q_t^h$	α_0^h (en %)	α_t^h (en %)
a	4	5	8	2,25	20	18	20	17
b	8	5	6	5	40	30	40	28
c	10	4	12	5	40	60	40	55
Total		14		12,25	100	108	100	100

1. Au regard des valeurs des indices élémentaires des prix figurant dans le tableau 3.8, on constate que sur la période 2010-2014, le prix du bien a a augmenté de 100 %, celui du bien b a diminué de 25 % et celui du bien c a augmenté de 20 %.

2. Les prix et les quantités des trois types de produits vendus en 2014 sont reportés dans le tableau 3.4 et sont obtenus comme suit. On sait que :

$$i_{t/0}^h(p) = \frac{p_t^h}{p_0^h} \times 100$$

et :

$$i^h_{t/0}(q) = \frac{q^h_t}{q^h_0} \times 100$$

On déduit donc de ces relations les expressions suivantes pour les prix et quantités en 2014 :

$$p^h_t = \frac{i^h_{t/0}(p)p^h_0}{100}$$

et :

$$q^h_t = \frac{i^h_{t/0}(q)q^h_0}{100}$$

3. Afin de comparer les structures des chiffres d'affaires en 2010 et en 2014 selon les trois types de produits, il convient de calculer les coefficients budgétaires α^h_0 et α^h_t :

$$\alpha^h_0 = \frac{p^h_0 q^h_0}{\Sigma_h p^h_0 q^h_0}$$

$$\alpha^h_t = \frac{p^h_t q^h_t}{\Sigma_h p^h_t q^h_t}$$

Au regard des valeurs figurant dans le tableau 3.4, on constate que les parts que représentent les produits a et b dans le chiffre d'affaires diminuent, alors que celle du produit c augmente.

4. Le nombre total de produits vendus est égal à 14 en 2010 et 12,25 en 2014. On en déduit la valeur de l'indice élémentaire de la quantité totale de produits vendus (base 1 en 2010) :

$$I_{t/0}(q) = \frac{12{,}25}{14} = 0{,}875$$

La quantité totale vendue a diminué de 12,5 % entre 2010 et 2014.

5. On sait que l'indice de valeur est donné par :

$$I^V_{t/0} = I_{t/0}(\bar{p}) \times I_{t/0}(q)$$

On en déduit la valeur de l'indice élémentaire du prix moyen $I_{t/0}(\bar{p})$ des produits considérés :

$$I_{t/0}(\bar{p}) = \frac{1{,}080}{0{,}875} = 1{,}234$$

Le prix moyen des produits vendus a augmenté de 23,4 % entre 2010 et 2014. Par ailleurs, le rapport $\frac{L_{t/0}(q)}{I_{t/0}(q)}$ correspond à l'indice de qualité S'_t. Cet indice étant égal à 1,133, on en conclut que la variation de la structure des ventes (effet qualité) vers les produits les plus chers a pour conséquence une augmentation du chiffre d'affaires de 13,3 %.

Chapitre 4

1 Schéma de décomposition additif et coefficients saisonniers

a. Faux.
b. Faux.
c. Faux.
d. Vrai, encore appelée série désaisonnalisée.
e. Faux.

2 Schéma de décomposition additif

a. Faux, c'est le cas dans un schéma multiplicatif.
b. Faux.
c. Vrai.
d. Faux.
e. Vrai.

3 Principe de conservation des aires et moyennes mobiles

a. Faux.
b. Vrai.
c. Vrai.
d. Faux.
e. Vrai.

4 Lissage exponentiel

a. Faux, le LES ne s'applique qu'au cas de séries avec tendance constante.
b. Faux, le LED ne s'applique pas au cas de séries présentant une composante saisonnière.
c. Faux, plus le paramètre de lissage est proche de 1, plus le poids des observations récentes est important.
d. Faux, le LES et le LED ne peuvent être appliqués si la série comporte une composante saisonnière.
e. Vrai.

5 Indice Euro Stoxx

Dans la mesure où les deux mois ne comportent pas le même nombre de jours, il faut calculer les valeurs corrigées des jours ouvrés. En considérant une moyenne de 21 jours par mois, les valeurs corrigées sont données par : $(3093{,}124/23) \times 21 = 2824{,}157$ en janvier et $(3085{,}865/20) \times 21 = 3240{,}158$ en février. On en déduit donc : $(3240{,}158 - 2824{,}157)/2824{,}157 = 0{,}1473$.

a. Faux.
b. Faux.
c. Faux.
d. Vrai.
e. Faux.

6 Étude de l'évolution de l'indice Euro Stoxx 50

Notons préalablement que les techniques de lissage doivent être appliquées à des séries dites stationnaires, c'est-à-dire à des séries dont la moyenne est stable au cours du temps. L'indice boursier étant ici considéré sur une période très courte (un mois), cette propriété est vérifiée.

1. L'application de la formule du LES :

$$L_t = \alpha Y_t + (1 - \alpha)L_{t-1}$$

nous permet d'obtenir les valeurs lissées L_t de la série Y_t ainsi qu'elles sont reportées dans le tableau 4.1 pour chacune des trois valeurs considérées du paramètre de lissage α. Nous avons retenu comme valeur de départ pour L_0 la première valeur de la série Y_1, c'est-à-dire la valeur observée au 01/04/2014.

▼ **Tableau 4.1** **Indice Euro Stoxx 50. Application du LES**

			$\alpha = 0{,}2$			$\alpha = 0{,}5$			$\alpha = 0{,}74$		
t	Date	Y_t	L_t	u_t	u_t^2	L_t	u_t	u_t^2	L_t	u_t	u_t^2
1	01/04/2014	3186,336	3186,336	0,000	0,000	3186,336	0,000	0,000	3186,336	0,000	0,000
2	02/04/2014	3187,450	3186,559	1,114	1,241	3186,893	1,114	1,241	3187,160	1,114	1,241
3	03/04/2014	3206,759	3190,599	20,200	408,048	3196,826	19,866	394,658	3201,663	19,599	384,107
4	04/04/2014	3230,332	3198,545	39,733	1578,724	3213,579	33,506	1122,652	3222,878	28,669	821,891
5	07/04/2014	3185,967	3196,030	–12,578	158,218	3199,773	–27,612	762,423	3195,564	–36,911	1362,433
6	08/04/2014	3177,658	3192,355	–18,372	337,522	3188,716	–22,115	489,073	3182,314	–17,906	320,621
7	09/04/2014	3182,793	3190,443	-9,562	91,440	3185,754	-5,923	35,076	3182,668	0,479	0,230
8	10/04/2014	3152,864	3182,927	–37,579	1412,177	3169,309	–32,890	1081,769	3160,613	–29,804	888,299
9	11/04/2014	3116,540	3169,650	-66,387	4407,254	3142,925	-52,769	2784,581	3127,999	–44,073	1942,441
10	14/04/2014	3131,566	3162,033	–38,084	1450,370	3137,245	-11,359	129,017	3130,639	3,567	12,723
11	15/04/2014	3091,524	3147,931	–70,509	4971,516	3114,385	-45,721	2090,436	3101,694	–39,115	1529,951
12	16/04/2014	3139,264	3146,198	-8,667	75,120	3126,824	24,879	618,983	3129,496	37,570	1411,521
13	17/04/2014	3155,806	3148,119	9,608	92,319	3141,315	28,982	839,938	3148,965	26,310	692,229
14	18/04/2014	3155,806	3149,657	7,687	59,084	3148,561	14,491	209,984	3154,027	6,841	46,795
15	21/04/2014	3155,806	3150,887	6,149	37,814	3152,183	7,245	52,496	3155,344	1,779	3,163
16	22/04/2014	3199,686	3160,646	48,799	2381,384	3175,935	47,503	2256,507	3188,157	44,342	1966,251
17	23/04/2014	3175,973	3163,712	15,327	234,903	3175,954	0,038	0,001	3179,141	–12,184	148,449
18	24/04/2014	3189,809	3168,931	26,097	681,066	3182,881	13,855	191,966	3187,035	10,668	113,810
19	25/04/2014	3147,397	3164,624	–21,534	463,722	3165,139	–35,484	1259,143	3157,703	–39,638	1571,193
20	28/04/2014	3165,837	3164,867	1,213	1,470	3165,488	0,698	0,487	3163,722	8,134	66,163
21	29/04/2014	3208,685	3173,631	43,818	1920,026	3187,087	43,197	1865,972	3196,995	44,963	2021,658
22	30/04/2014	3198,387	3178,582	24,756	612,883	3192,737	11,300	127,700	3198,025	1,392	1,939
Somme					21376,30			16314,10			15307,11

2. La valeur prévue $\hat{Y}_t$ de Y_t est donnée par :

$$\hat{Y}_t = L_t$$

et l'erreur de prévision u_t est donnée par :

$$u_t = Y_t - \hat{d}_{t-1} = Y_t - L_{t-1}$$

On déduit, d'après les résultats reportés dans le tableau 4.1, les valeurs de la somme des carrés des erreurs de prévision :

– Pour $\alpha = 0{,}2$, on a : $\Sigma u_t^2 = 21376{,}30$

– Pour $\alpha = 0{,}5$, on a : $\Sigma u_t^2 = 16314{,}10$

– Pour $\alpha = 0{,}74$, on a : $\Sigma u_t^2 = 15307{,}11$

3. Si l'on retient pour α la valeur qui minimise la somme des carrés des erreurs de prévision, il convient de choisir la valeur $\alpha = 0{,}74$. Ce résultat, qui consiste à retenir une valeur relativement élevée de α, n'est pas surprenant dans la mesure où nous étudions une série boursière, qui constitue un exemple typique de série relativement heurtée.

4. La valeur prévue de la série pour le 01/05/2014, c'est-à-dire pour $t = 23$ est donnée par la dernière valeur obtenue pour la série lissée :

$$\hat{Y}_{23} = L_{22}$$

soit 3198,025 pour une valeur du paramètre de lissage égale à 0,74. Dans la mesure où la prévision est indépendante de l'horizon dans le cas du LES, on a :

$$\hat{Y}_{24} = L_{22}$$

et la valeur prévue au 30/04/2014 de la série pour le 02/05/2014 est donc égale à 3198,025 pour $\alpha = 0{,}74$.

7 Étude des ventes de voitures neuves en France

1. Le tableau 4.2 est un tableau à double entrée dans lequel on a reporté en ligne les années et en colonne les trimestres. Ce tableau est appelé tableau de Buys-Ballot.

2. Il est possible de classer, pour chaque année, les trimestres en fonction des valeurs décroissantes des ventes de voitures. Le tableau 4.3 reporte les trimestres ainsi classés, le chiffre figurant dans chaque case correspondant au numéro du trimestre. On constate que le deuxième trimestre est en général classé en première position alors que le troisième trimestre est systématiquement classé en dernière position. Cela témoigne d'un pic dans les ventes de voitures durant le deuxième trimestre de chaque année et d'un creux au cours du troisième trimestre, indiquant l'existence d'une saisonnalité dans la série des ventes de voitures neuves en France. Ce résultat est cohérent avec ce que l'on observe sur le marché automobile français puisque les ventes de voitures neuves sont en effet caractérisées par une forte hausse au mois de juin et un creux durant le mois d'août.

3. Les valeurs demandées de la moyenne et de l'écart-type sont reportées dans les deux dernières colonnes du tableau de Buys-Ballot (tableau 4.2).

4. Notons $\bar{Y}_i$ la moyenne des ventes de voitures pour chaque année i et σ_i l'écart-type de cette même série. Le coefficient a de la droite de régression de σ_i sur $\bar{Y}_i$ est donné par :

$$a = \frac{Cov(\bar{Y}_i, \sigma_i)}{V(\bar{Y}_i)} = \frac{340\,638\,745}{1\,651\,883\,555} = 0{,}2062$$

Le coefficient de la pente de la droite de régression étant significativement différent de zéro, cela témoigne de l'existence d'un lien entre la moyenne et l'écart-type des ventes de voitures. Cela peut être confirmé par le calcul du coefficient de corrélation entre les deux séries :

$$r(\bar{Y}_i, \sigma_i) = \frac{Cov(\bar{Y}_i, \sigma_i)}{\sigma_{\bar{Y}_i} \times \sigma_{\sigma_i}}$$

$$= \frac{340\,638\,745}{40\,643{,}3704 \times 11\,497{,}1203} = 0{,}7\,290$$

L'existence d'un lien entre les valeurs annuelles de la moyenne et celles de l'écart-type conduit à retenir un schéma de décomposition multiplicatif.

▼ Tableau 4.2 Tableau de Buys-Ballot

	Trimestre 1	Trimestre 2	Trimestre 3	Trimestre 4	Moyenne	Écart-type
2006	526 502	582 727	409 138	482 182	500 137,25	63 481,14
2007	519 191	561 417	442 753	541 182	516 135,75	44 922,35
2008	526 121	602 775	446 869	474 518	512 570,75	59 340,43
2009	505 456	625 859	482 186	655 170	567 167,75	74 530,69
2010	590 869	597 383	432 029	589 905	552 546,5	69 640,24
2011	642 627	557 981	422 728	537 592	540 232	78 444,12
2012	501 879	520 836	370 006	464 292	464 253,25	58 094,43
2013	431 388	480 730	367 112	477 721	439 237,75	46 006,28
Moyenne	530 504,13	566 213,50	421 602,63	527 820,25		
Écart-type	58 789,33	44 152,99	36 491,19	63 063,71		

▼ Tableau 4.3 Classement des trimestres en fonction des valeurs décroissantes des ventes

Année	T1	T2	T3	T4
2006	2	1	4	3
2007	2	4	1	3
2008	2	1	4	3
2009	4	2	1	3
2010	2	1	4	3
2011	1	2	4	3
2012	2	1	4	3
2013	2	4	1	3

Chapitre 5

1 Expérience aléatoire

a. Vrai.
b. Vrai.
c. Vrai.
d. Faux. L'univers des possibles se compose d'événements élémentaires, mais aussi d'événements composés (union ou intersection de singletons).
e. Faux. L'univers des possibles peut être fini ou infini (dénombrable ou non dénombrable).

2 Événement

a. Vrai.
b. Vrai.
c. Faux. L'intersection de deux événements disjoints est l'ensemble vide.
d. Faux. L'union d'un événement et son complémentaire ne correspond pas nécessairement à l'univers des possibles. Elle peut correspondre à un autre événement.
e. Vrai. On peut toujours considérer que l'ensemble vide fait partie d'un ensemble d'événements quelconque.

3 Ensemble des parties et tribu

a. Faux. Ceci n'est vrai que dans le cas particulier où l'univers des possibles ne comprend qu'un seul résultat.
b. Vrai.
c. Vrai.
d. Vrai. C'est la propriété de stabilité par passage au complémentaire
e. Faux. D'après la propriété de stabilité par réunion dénombrable, c'est l'union d'événements disjoints appartenant à une tribu, qui appartient elle aussi à cette tribu.

4 Mesure de probabilité

a. Faux. Une mesure de probabilité peut être définie sur un univers probabilisable.
b. Vrai.
c. Vrai. Dans le cas d'un univers fini ou dénombrable, l'ensemble des parties est une tribu. L'univers et l'ensemble des parties forment alors un univers probabilisable.
d. Vrai.
e. Faux. Si $B \subset A$, alors $\Pr(B) \leq \Pr(A)$.

5 Probabilité et dénombrement

1. Une grille-réponses est une suite de 4 réponses. Pour chaque question il y a 5 réponses possibles. Il y a donc $5^4 = 625$ grilles-réponses possibles.

2. L'événement A « répondre au hasard correctement à au moins 2 questions » est réalisé si l'étudiant répond correctement à 2, 3 ou 4 questions. Notons A_i pour $i \in \{0,1,2,3,4\}$ l'événement « répondre au hasard correctement à i questions ». On a donc $A = A_2 \cup A_3 \cup A_4$. Si l'étudiant répond au hasard, la probabilité associée à l'événement A_i est égale à :

$$\Pr(A_i) = \frac{C_4^i \times 4^{4-i}}{5^4}$$

En effet, si l'étudiant répond correctement à i questions parmi 4, il a $4-i$ questions fausses. Pour chaque question fausse, il 4 réponses (fausses) possibles sur les 5 choix proposés. Ainsi, on obtient

$$\Pr(A_4) = \frac{C_4^4 \times 4^{4-4}}{5^4} = \frac{1}{625}$$
$$\Pr(A_3) = \frac{C_4^3 \times 4^{4-3}}{5^4} = \frac{16}{625}$$
$$\Pr(A_2) = \frac{C_4^2 \times 4^{4-2}}{5^4} = \frac{96}{625}$$

Ainsi, la probabilité que l'étudiant obtienne au moins la moyenne sur cet exercice est égale à environ 18 %.

$$\Pr(A) = \Pr(A_2) + \Pr(A_3) + \Pr(A_4) = \frac{113}{625} = 0{,}1808$$

6 Probabilité et dénombrement

L'univers des possibles Ω correspond à l'ensemble des combinaisons de 4 billets parmi les 50. Il y en a :

$$card\,(\Omega) = C_{50}^4 = \binom{50}{4} = \frac{50!}{(50-4)! \times 4!} = 230\,300$$

On ne gagne rien si les 4 billets achetés figurent parmi les 48 billets perdants. La probabilité de ne rien gagner, événement noté A, est donc égale à :

$$\Pr(A) = \frac{C_{48}^4}{C_{50}^4} = \frac{1}{230\,300} \times \frac{48!}{(48-4)! \times 4!} = \frac{194\,580}{230\,300} = 0{,}8449$$

L'événement « gagner au moins un lot » correspond à l'événement complémentaire de l'événement « ne rien gagner ». Par conséquent, la probabilité associée est égale à :

$$\Pr\left(\overline{A}\right) = 1 - 0{,}8449 = 0{,}1551$$

Il y a 15,51 % de chances de gagner au moins un lot dans cette loterie.

7 Suite d'événements et probabilité totale

1. On note F_n l'événement « fumer le jour n » et $\overline{F}_n$ l'événement complémentaire « ne pas fumer le jour n ». L'ensemble d'événements $\left\{F_n, \overline{F}_n\right\}$ est un système complet, on peut donc appliquer la formule des probabilités totales pour déterminer la probabilité le jour $n+1$:

$$\begin{aligned}\Pr(F_{n+1}) &= \Pr(F_{n+1}|F_n) \times \Pr(F_n) + \Pr\left(F_{n+1}|\overline{F}_n\right) \\ &\quad \times \Pr\left(\overline{F}_n\right) \\ &= 0{,}8 \times \Pr(F_n) + 0{,}1 \times (1 - \Pr(F_n))\end{aligned}$$

On obtient donc une formule de récurrence du type :

$$\Pr(F_{n+1}) = 0{,}7 \times \Pr(F_n) + 0{,}1$$

2. On considère la suite $P_{n+1} = 0{,}7P_n + 0{,}1$. Une solution particulière est donnée par la solution de l'équation :

$$x = 0{,}7x + 0{,}1 \Longleftrightarrow x = \frac{1}{3}$$

La solution générale est la solution de $P_{n+1} = 0{,}7P_n$, c'est-à-dire $P_n = 0{,}7^{n-1}P_1$ où P_1 désigne la probabilité fumer le premier jour. La probabilité de fumer le jour n est donc donnée par :

$$\Pr(F_n) = 0{,}7^{n-1} \times \Pr(F_1) + \frac{1}{3}$$

Lorsque $n \to \infty$, on obtient $\Pr(F_n) = 1/3$. Il y a donc 66,66 % de chances que cette personne finisse par s'arrêter de fumer.

8 Probabilité et indépendance

1. Pour un individu de la population, la probabilité d'achat des deux produits est égale à :

$$\Pr(A \cap B) = p_A \times p_B$$

car les ventes des deux produits sont indépendantes.

2. Pour un individu de la population, la probabilité d'achat de l'un ou de l'autre produit est égale à :

$$\begin{aligned}\Pr(A \cup B) &= \Pr(A) + \Pr(B) - \Pr(A \cap B)\\ &= p_A + p_B - p_A \times p_B\end{aligned}$$

9 Probabilité et dénombrement

1. Soit $p = N_R/N$ la probabilité de tirer une boule rouge. À chaque tirage il y a deux éventualités : soit tirer une boule rouge (événement R), soit tirer une boule blanche (événement B) avec $\Pr(R) = p$ et $\Pr(B) = 1 - p$ ce qui définit un système complet sur l'univers des possibilités $\Omega = \{B,R\}$. Faire n tirages avec remise revient à considérer l'espace produit probabilisable $\Omega^n = \{(x_1,\ldots,x_n)\,;\, x_i = B \text{ ou } R\ \forall i\}$ muni de la probabilité p^n. Soit l'événement E_k « on tire k boules rouges » composé des éléments $\omega = (x_1,\ldots,x_n)$ tels que parmi les singletons x_i il y k fois la lettre R et $n - k$ fois la lettre B. Il y a donc C_n^k événements correspondants, c'est-à-dire C_n^k façons de disposer k boules rouges à n places numérotées. Chacune de ces éventualités a pour probabilité $p^k(1-p)^{n-k}$ car les tirages sont indépendants.

$$P(E_k) = C_n^k \left(\frac{N_R}{N}\right)^k \left(1 - \frac{N_R}{N}\right)^{n-k}$$

2. On vérifie que la probabilité de tirer $k = 10$ boules rouges est égale à 9,85 %, puisque :

$$P(E_{10}) = C_{50}^{10} \left(\frac{25}{100}\right)^{10} \left(1 - \frac{25}{100}\right)^{50-10} = 9{,}85\ \%$$

10 Probabilité conditionnelle

On note D l'événement « lot défecteux » et $\overline{D}$ l'événement « lot valide ». On note T l'événement « le test conduit au rejet du lot » et $\overline{T}$ l'événement « le test ne conduit pas au rejet du lot ».

1. La probabilité qu'un lot soit effectivement défectueux si le test conduit au rejet du lot est égale à :

$$\Pr(D|T) = \frac{\Pr(D \cap T)}{\Pr(T)} = \frac{\Pr(T|D) \times \Pr(D)}{\Pr(T)}$$

Or la probabilité qu'un test conduise au rejet d'un lot est égale à :

$$\begin{aligned}\Pr(T) &= \Pr(T|D) \times \Pr(D) + \Pr\left(T|\overline{D}\right) \times \Pr\left(\overline{D}\right)\\ &= 0{,}98 \times 0{,}05 + 0{,}04 \times 0{,}95 = 0{,}087\end{aligned}$$

On en déduit que :

$$\Pr(D|T) = \frac{\Pr(T|D) \times \Pr(D)}{\Pr(T)} = \frac{0{,}98 \times 0{,}05}{0{,}087} = 0{,}5632$$

2. La probabilité qu'un lot soit valide si le test conduit au rejet du lot est égale à la probabilité complémentaire :

$$\Pr\left(\overline{D}|T\right) = 1 - \Pr(D|T) = 1 - 0{,}5632 = 0{,}4368$$

3. La probabilité qu'un lot soit défectueux si le test ne conduit pas au rejet du lot est égale à :

$$\begin{aligned}\Pr\left(D|\overline{T}\right) &= \frac{\Pr\left(D \cap \overline{T}\right)}{\Pr\left(\overline{T}\right)} = \frac{\Pr\left(\overline{T}|D\right) \times \Pr(D)}{\Pr\left(\overline{T}\right)}\\ &= \frac{0{,}02 \times 0{,}05}{1 - 0{,}087} = 0{,}0011\end{aligned}$$

4. La probabilité qu'un lot soit valide si le test ne conduit pas au rejet du lot est égale à la probabilité complémentaire :

$$\Pr\left(\overline{D}|\overline{T}\right) = 1 - \Pr\left(D|\overline{T}\right) = 1 - 0{,}0011 = 0{,}9989$$

11 Probabilité conditionnelle

Soit A l'événement « l'étudiant est issu de la formation A » et B l'événement « l'étudiant est issu de la formation B ». Soit M l'événement « l'étudiant a obtenu une mention bien ». La probabilité qu'un étudiant ayant obtenu une mention bien soit issu de la formation A est égale à :

$$\Pr(A|M) = \frac{\Pr(A \cap M)}{\Pr(M)} = \frac{\Pr(M|A) \times \Pr(A)}{\Pr(M)}$$

La probabilité d'obtenir une mention est égale à :

$$\begin{aligned}\Pr(M) &= \Pr(M|A) \times \Pr(A) + \Pr(M|B) \times \Pr(B)\\ &= 0{,}30 \times 0{,}70 + 0{,}20 \times 0{,}30 = 0{,}27\end{aligned}$$

Par conséquent :

$$\Pr(A|M) = \frac{0{,}30 \times 0{,}70}{0{,}27} = 0{,}7778$$

12 Probabilité conditionnelle

1. Nous avons :

$$\begin{aligned}\Pr(A_{n+1}) &= \Pr(A_{n+1}|A_n) \times \Pr(A_n) + \Pr\left(A_{n+1}|\overline{A}_n\right)\\ &\quad \times \Pr\left(\overline{A}_n\right)\\ &= \frac{1}{8}\Pr(A_n) + \frac{3}{8}(1 - \Pr(A_n))\\ &= \frac{3}{8} - \frac{2}{8}\Pr(A_n)\end{aligned}$$

2. La relation précédente définit une suite géométrique. Une solution particulière est donnée par

$$x = \frac{3}{8} - \frac{2}{8}x \Longleftrightarrow x = \frac{3}{10}$$

On en déduit que la probabilité de l'événement « l'individu se rend au cinéma le jour i » est égale à :

$$\Pr(A_n) = \frac{3}{10} + p_1\left(-\frac{2}{8}\right)^{n-1}$$

avec $\Pr(A_1) = p_1$.

Chapitre 6

1 Variable aléatoire

a. Faux. Une variable aléatoire est une application mesurable.
b. Faux. Seules les variables quantitatives peuvent être continues.
c. Vrai.
d. Vrai. Une variable aléatoire continue est définie sur un support infini non dénombrable.
e. Faux. Pas nécessairement, une variable aléatoire discrète peut être définie sur un support de dimension finie.

2 Fonction de densité, de masse et de répartition

a. Vrai.
b. Faux. La densité est toujours positive, mais elle peut être supérieure à 1 en un point.
c. Faux. La fonction de densité est la dérivée de la fonction de répartition.
d. Vrai. C'est la définition de la fonction de répartition.
e. Faux. La fonction de masse ne correspond pas à la dérivée de la fonction de répartition.

3 Fonction de répartition et quantile

a. Faux. Un quantile est une réalisation.
b. Vrai.
c. Faux. Les quantiles sont définis sur le support de la loi de probabilité.
d. Faux. Pour de nombreuses lois usuelles, la fonction de répartition n'a pas d'expression analytique.
e. Vrai.

4 Indépendance

a. Faux. La covariance n'est qu'une mesure de dépendance linéaire. Il peut exister d'autres formes de dépendance non-linéaire entre ces variables.
b. Faux. Si la densité jointe est égale au produit des densités marginales, les variables sont indépendantes.
c. Vrai.
d. Vrai.
e. Vrai.

5 Fonction de répartition

Soit X une variable aléatoire réelle admettant pour fonction de répartition :

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1/4 & \text{si } 0 \leq x < 1 \\ 3/4 & \text{si } 1 \leq x < 2 \\ 1 & \text{si } x \geq 2 \end{cases}$$

1.

$$\begin{aligned} \Pr\left(-\frac{1}{2} < X < \frac{1}{2}\right) &= \Pr\left(X < \frac{1}{2}\right) - \Pr\left(X \leq -\frac{1}{2}\right) \\ &= F_X\left(\frac{1}{2}\right) - F_X\left(-\frac{1}{2}\right) \\ &= \frac{1}{4} - 0 = \frac{1}{4} \end{aligned}$$

2.

$$\begin{aligned} \Pr\left(-\frac{1}{2} < X < \frac{3}{2}\right) &= \Pr\left(X < \frac{3}{2}\right) - \Pr\left(X \leq -\frac{1}{2}\right) \\ &= F_X\left(\frac{3}{2}\right) - F_X\left(-\frac{1}{2}\right) \\ &= \frac{3}{4} - 0 = \frac{3}{4} \end{aligned}$$

3.

$$\begin{aligned} \Pr(X > 3) &= 1 - \Pr(X \leq 3) \\ &= 1 - F_X(3) \\ &= 1 - 1 \\ &= 0 \end{aligned}$$

6 Espérance

1. Par définition :

$$\begin{aligned} \mathbb{E}((X-a)^2) &= \mathbb{V}(X-a) + (\mathbb{E}(X-a))^2 \\ &= \mathbb{V}(X) + (\mathbb{E}(X-a))^2 \end{aligned}$$

2. On pose $f(a) = \mathbb{E}((X-a)^2)$. Cette fonction atteint son minimum lorsque $\partial f(a)/\partial a = 0$ et $\partial^2 f(a)/\partial a^2 > 0$.

$$\frac{\partial f(a)}{\partial a} = \mathbb{E}(-2(X-a)) = \mathbb{E}(-2X+2a) = -2\mathbb{E}(X)+2a = 0$$

$$\frac{\partial^2 f(a)}{\partial a^2} = 2 > 0$$

On en déduit que $a = \mathbb{E}(X)$ est un minimum. Donc la valeur de a pour laquelle l'espérance $\mathbb{E}\left((X-a)^2\right)$ est minimum est égale à $\mathbb{E}(X)$.

7 Fonction de densité

1. La fonction $f_X(x)$ définie par :

$$f_X(x) = \begin{cases} 0 & \text{si } x \leq \theta \\ \exp(-(x-\theta)) & \text{si } x > \theta \end{cases}$$

est positive $\forall x \in \mathbb{R}$. Montrons qu'elle intègre à l'unité sur le support $X(\Omega) = \mathbb{R}$.

$$\begin{aligned} \int_{-\infty}^{+\infty} f_X(x)\,dx &= \int_{\theta}^{+\infty} \exp(-(x-\theta))\,dx \\ &= \int_0^{+\infty} \exp(-x)\,dx \\ &= \Gamma(1) = 0! = 1 \end{aligned}$$

où $\Gamma(.)$ désigne la fonction gamma. Donc $f_X(x)$ satisfait les conditions d'une fonction de densité.

2. Par définition de la fonction de répartition, $\forall x \in \mathbb{R}$:

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(u)\,du = \int_{\theta}^{x} \exp(-(u-\theta))\,du \\ &= \exp(\theta) \int_{\theta}^{x} \exp(-u)\,du \\ &= \exp(\theta) \times [-\exp(-u)]_{\theta}^{x} \\ &= 1 - \exp(-(x-\theta)) \end{aligned}$$

3. Par définition, l'espérance $\mathbb{E}(X)$ est égale à :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f_X(x)\,dx = \int_{\theta}^{+\infty} x \exp(-(x-\theta))\,dx$$

En posant $u = x - \theta \Longleftrightarrow x = u + \theta$, $du = dx$ il vient :

$$\begin{aligned}\mathbb{E}(X) &= \int_{\theta}^{+\infty} (u+\theta)\exp(-u)\,du \\ &= \int_{\theta}^{+\infty} u\exp(-u)\,du + \int_{\theta}^{+\infty} \theta\exp(-u)\,du \\ &= \Gamma(2) + \theta\Gamma(1) = 1! + \theta \times 0! = 1 + \theta\end{aligned}$$

4. De la même façon :

$$\mathbb{E}\left(X^2\right) = \int_{-\infty}^{+\infty} x^2 f_X(x)\,dx = \int_{\theta}^{+\infty} x^2 \exp(-(x-\theta))\,dx$$

En posant $u = x - \theta \Longleftrightarrow x = u + \theta$, $du = dx$ il vient :

$$\begin{aligned}\mathbb{E}\left(X^2\right) &= \int_{\theta}^{+\infty} (u+\theta)^2\exp(-u)\,du \\ &= \int_{\theta}^{+\infty} \left(u^2 + 2u\theta + \theta^2\right)\exp(-u)\,du \\ &= \int_{\theta}^{+\infty} u^2\exp(-u)\,du + 2\theta\int_{\theta}^{+\infty} u\exp(-u)\,du + \theta^2 \\ &\quad \times \int_{\theta}^{+\infty} \exp(-u)\,du \\ &= \Gamma(3) + 2\theta\Gamma(2) + \theta^2\Gamma(1) \\ &= \theta^2 + 2\theta + 2\end{aligned}$$

On en déduit la variance :

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2 \\ &= \theta^2 + 2\theta + 2 - (1+\theta)^2 \\ &= \theta^2 + 2\theta + 2 - 1 - 2\theta - \theta^2 = 1\end{aligned}$$

8 Transformée de variable aléatoire

1. Puisque les variables X et Y sont indépendantes :

$$\Pr(Z > k) = \Pr((X>k)\cap(Y>k)) = \Pr(X>k) \times \Pr(Y>k)$$

Par ailleurs, nous savons que :

$$\begin{aligned}\Pr(X > k) &= \Pr(X \geq k+1) \\ &= \Pr(X = k+1) + \Pr(X = k+2) + \ldots \\ &= \sum_{i=k+1}^{+\infty} \Pr(X = i) \\ &= \sum_{i=k+1}^{+\infty} q^{i-1}p\end{aligned}$$

En posant $j = i - k - 1$, on obtient :

$$\Pr(X > k) = \sum_{j=0}^{+\infty} q^{j+k}p = q^k p \sum_{j=0}^{+\infty} q^j = \frac{q^k p}{1-q} = q^k$$

De la même façon $\Pr(Y > k) = q^k$. On en conclut que :

$$\Pr(Z > k) = q^{2k} \quad \forall k \in \mathbb{N}^*$$

2. Par définition :

$$\Pr(Z \geq k) = \Pr((Z > k) \cup (Z = k))$$

Puisque les deux événements $(Z > k)$ et $(Z = k)$ sont incompatibles, on a :

$$\Pr(Z \geq k) = \Pr(Z > k) + \Pr(Z = k)$$

ou encore :

$$\Pr(Z = k) = \Pr(Z \geq k) - \Pr(Z > k)$$

Or puisque Z est définie sur $\mathbb{N}^*$, on a $\Pr(Z \geq k) = \Pr(Z > k-1)$. Par conséquent :

$$\Pr(Z = k) = \Pr(Z > k-1) - \Pr(z > k)$$

3. D'après les résultats précédents, on montre que :

$$\begin{aligned}\Pr(Z = k) &= \Pr(Z > k-1) - \Pr(z > k) \\ &= q^{2(k-1)} - q^{2k} \\ &= q^{2k-2}\left(1 - q^2\right) \\ &= \left(q^2\right)^{k-1}\left(1 - q^2\right)\end{aligned}$$

Donc Z suit une loi géométrique de paramètre $\left(1 - q^2\right)$.

9 Variable aléatoire discrète

1. Au premier tour, il y a 2 rois rouges dans le paquet de $2n$ cartes. La probabilité de l'événement E_1 « le premier roi rouge obtenu est la 1re carte retournée » est donc égale à :

$$\Pr(E_1) = \frac{2}{2n}$$

Au second tour, la probabilité conditionnelle à l'échec du premier tour, c'est-à-dire à $\overline{E}_1$, de tirer un roi rouge parmi les $2n - 1$ cartes restantes est égale à :

$$\Pr\left(E_2|\overline{E}_1\right) = \frac{2}{2n-1}$$

La probabilité d'avoir échoué au premier tirage est égale à :

$$\Pr\left(\overline{E}_1\right) = 1 - \Pr(E_1) = \frac{2n-2}{2n}$$

Dès lors, la probabilité de l'événement E_2 est égale à :

$$\begin{aligned}\Pr(E_2) &= \Pr\left(E_2|\overline{E}_1\right) \times \Pr\left(\overline{E}_1\right) \\ &= \frac{2}{2n-1} \times \frac{2n-2}{2n} = \frac{2n-2}{n(2n-1)}\end{aligned}$$

Au troisième tour, la probabilité de l'événement E_3 peut s'écrire sous la forme :

$$\begin{aligned}\Pr(E_3) &= \Pr\left(E_3|\overline{E}_2,\overline{E}_1\right) \times \Pr\left(\overline{E}_2\right) \times \Pr\left(\overline{E}_1\right) \\ &= \frac{2}{2n-2} \times \frac{2n-3}{2n-1} \times \frac{2n-2}{2n} = \frac{2n-3}{n(2n-1)}\end{aligned}$$

De façon générale, on montre que :

$$\begin{aligned}\Pr(E_k) &= \Pr\left(E_k|\overline{E}_{k-1},\ldots,\overline{E}_1\right) \times \Pr\left(\overline{E}_{k-1}\right) \times \ldots \times \Pr\left(\overline{E}_1\right) \\ &= \frac{2n-k}{n(2n-1)}\end{aligned}$$

On vérifie que la probabilité de tirer le premier roi rouge au $k^{ème}$ tour est nulle, puisque cet événement est impossible sachant qu'il y a deux rois rouges dans le jeu de cartes, et que nécessairement le premier roi rouge a été tiré entre le premier tour et le $k-1^{ème}$ tour.

$$\Pr(E_{2n}) = \frac{2n-2n}{n(2n-1)} = 0$$

2. Soit $E(\Omega) = \{E_1,\ldots,E_{2n}\}$ l'univers des résultats possibles. On définit une variable aléatoire X correspondant au gain : à chaque tour le joueur perd 1 euro et il gagne a euros s'il tire un roi rouge. S'il tire le roi au 1er tour, il gagne $a-1$ euros, s'il tire le roi au deuxième tour il gagne $a-2$ euros. Ainsi, s'il tire un roi rouge au $k^{ème}$ tour, il gagne $a-k$ euros. La variable X est donc définie sur le support $X(\Omega) = \{a-1, a-2, \ldots, a-2n\}$.

3. Par définition :

$$\Pr(X = a-k) = \Pr(E_k) = \frac{2n-k}{n(2n-1)}$$

4. On vérifie que :

$$\begin{aligned}\sum_{k=1}^{2n} \Pr(X=a-k) &= \frac{1}{n(2n-1)}\sum_{k=1}^{2n}(2n-k)\\ &= \frac{1}{n(2n-1)}\left(4n^2 - \frac{2n(2n+1)}{2}\right)\\ &= \frac{1}{n(2n-1)}\left(\frac{4n^2-2n}{2}\right)\\ &= 1\end{aligned}$$

Le système défini par les réalisations $X(\Omega)$ est un système complet.

10 Variable aléatoire continue

1. Pour que la fonction $f_X(x)$ soit une fonction de densité, elle doit être positive $\forall x \in \mathbb{R}$ et son intégrale sur $\mathbb{R}$ doit être égale à 1. La positivité est assurée dès lors que $a > 0$. Par ailleurs :

$$\begin{aligned}\int_{-\infty}^{+\infty} f_X(x)\,dx &= a\int_{-\infty}^{+\infty}\exp(-|x|)\,dx\\ &= 2a\int_0^{+\infty}\exp(-x)\,dx = 1\end{aligned}$$

On doit donc avoir $a = 1/2$. La fonction $f_X(x)$ est une densité si :

$$f_X(x) = \frac{1}{2}\exp(-|x|)$$

2. La fonction de répartition est définie par :

$$F_X(x) = \Pr(X \le x) = \frac{1}{2}\int_{-\infty}^{x}\exp(-|u|)\,du \quad \forall x \in \mathbb{R}$$

Lorsque $x \le 0$, on a :

$$\begin{aligned}F_X(x) &= \frac{1}{2}\int_{-\infty}^{x}\exp(-|u|)\,du = \frac{1}{2}\int_{-\infty}^{x}\exp(u)\,du\\ &= \frac{1}{2}[\exp(u)]_{-\infty}^{x} = \frac{\exp(x)}{2}\end{aligned}$$

Lorsque $x > 0$, on a :

$$\begin{aligned}F_X(x) &= \frac{1}{2}\int_{-\infty}^{x}\exp(-|u|)\,du\\ &= \frac{1}{2}\int_{-\infty}^{0}\exp(u)\,du + \frac{1}{2}\int_0^x\exp(-u)\,du\\ &= \frac{1}{2}[\exp(u)]_{-\infty}^{0} + \frac{1}{2}[-\exp(-u)]_0^x\\ &= \frac{1}{2} - \frac{\exp(-x)}{2} + \frac{1}{2} = 1 - \frac{\exp(-x)}{2}\end{aligned}$$

On obtient au final :

$$F_X(x) = \begin{cases}\dfrac{\exp(x)}{2} & \text{si } x \le 0\\ 1 - \dfrac{\exp(-x)}{2} & \text{si } x > 0\end{cases}$$

3. Par définition :

$$\begin{aligned}\mathbb{E}(X) &= \frac{1}{2}\int_{-\infty}^{+\infty} x\exp(-|x|)\,dx\\ &= \frac{1}{2}\int_{-\infty}^{0} x\exp(x)\,dx + \frac{1}{2}\int_0^{+\infty} x\exp(-x)\,dx\end{aligned}$$

En posant un changement de variable $z = -x$ sur la première intégrale, il vient :

$$\mathbb{E}(X) = -\frac{1}{2}\int_0^{+\infty} z\exp(-z)\,dz + \frac{1}{2}\int_0^{+\infty} x\exp(-x)\,dx = 0$$

Chapitre 7

1 Loi usuelles discrètes

a. Faux. La loi binomiale correspond au nombre de succès obtenus dans n tirages successifs et indépendants d'une expérience de Bernoulli.
b. Vrai.
c. Vrai.
d. Vrai.
e. Vrai.

2 Loi usuelles continues

a. Vrai.
b. Faux. Cette propriété n'est vraie que pour la loi exponentielle.
c. Faux. Sa skewness est positive.
d. Faux. Ceci n'est vrai que si $\mu = 0$ et $\sigma^2 = 1$.
e. Vrai.

3 Loi normale

a. Vrai.
b. Faux. $\Phi^{-1}(0{,}025) = -1{,}96$. De façon générale $\Phi^{-1}(\alpha) = -\Phi^{-1}(1-\alpha)$.
c. Vrai. La fonction de répartition inverse de la loi normale centrée réduite $\Phi^{-1}(\alpha)$ est croissante sur $[0,1]$.
d. Vrai. Puisque la fonction de densité de la loi normale centrée réduite est symétrique par rapport à 0, on a $\Phi(x) = -\Phi(-x)$.
e. Vrai.

4 Loi de Student

a. Faux. La kurtosis peut être égale à 3 si le nombre de degrés de liberté de la loi de Student tend vers l'infini.

b. Vrai.

c. Faux. Puisque la fonction de densité de la loi de Student est symétrique par rapport à 0, on a $F_X(x) = -F_X(-x)$.

d. Vrai.

e. Faux. Car si $\alpha < 0{,}5$ le quantile d'ordre α est nécessairement négatif. Si $X \sim t(4)$, le vrai quantile est $F_X^{-1}(0{,}05) = -2{,}1318$.

5 Loi usuelles discrètes

1. Par définition $\Pr(X = 4) = 0$ car $4 \notin X(\Omega)$.

$$\Pr(X = 1) = \frac{1}{3-(-2)+1} = \frac{1}{6}$$

$$\Pr(X \leq 2) = F_X(2) = \frac{2-(-2)+1}{3-(-2)+1} = \frac{5}{6}$$

Ou de façon équivalente :

$$\Pr(X \leq 2) = \sum_{i=-2}^{2} \Pr(X = i) = \sum_{i=-2}^{2} \frac{1}{6} = 5 \times \frac{1}{6} = \frac{5}{6}$$

2. Si $X \sim \mathcal{B}(10; 0{,}4)$ alors son support est $X(\Omega) = \{0, 1, \ldots, 10\}$. La réalisation $x = 2{,}57$ n'appartient pas à $X(\Omega)$ donc $\Pr(X = 2{,}57) = 0$.

$$\Pr(X = 2) = \binom{10}{2} \times 0{,}4^2 \times 0{,}6^{10-2} = 0{,}1209$$

Soit $F_X(x)$ la fonction de répartition de la loi $\mathcal{B}(10; 0{,}4)$:

$$\begin{aligned}\Pr(2{,}57 < X \leq 4) &= \Pr(X \leq 4) - \Pr(X \leq 2{,}57) \\ &= F_X(4) - F_X(2{,}57)\end{aligned}$$

Nous savons que :

$$\begin{aligned}F_X(2{,}57) &= \sum_{k=0}^{\lfloor 2{,}57 \rfloor} \binom{10}{k} \times 0{,}4^k \times (1-0{,}4)^{10-k} \\ &= \sum_{k=0}^{2} \binom{10}{k} \times 0{,}4^k \times (1-0{,}4)^{10-k} \\ &= F_X(2)\end{aligned}$$

Pour une loi $\mathcal{B}(10; 0{,}4)$, on lit dans la table fournie en annexe que :

$$F_X(4) = \Pr(X \leq 4) = 0{,}6331$$
$$F_X(2) = \Pr(X \leq 2) = 0{,}1673$$

On en déduit que :

$$\Pr(2{,}57 < X \leq 4) = 0{,}6331 - 0{,}1673 = 0{,}4658$$

3. Si $X \sim \mathcal{P}(4)$ alors son support est $X(\Omega) = \mathbb{N}$. La réalisation $x = 2{,}57$ n'appartient pas à $X(\Omega)$ donc $\Pr(X = 2{,}57) = 0$.

$$\Pr(X = 5) = \frac{4^5 \times \exp(-4)}{5!} = 0{,}1563$$

Soit $F_X(x)$ la fonction de répartition de la loi $\mathcal{P}(4)$:

$$\Pr(2{,}57 < X \leq 4) = \Pr(X \leq 4) - \Pr(X \leq 2{,}57) = F_X(4) - F_X(2{,}57)$$

Par définition :

$$F_X(2{,}57) = \exp(-4) \sum_{i=0}^{\lfloor 2{,}57 \rfloor} \frac{4^i}{i!} = \exp(-4) \sum_{i=0}^{2} \frac{4^i}{i!} = F_X(2)$$

Pour une loi $\mathcal{P}(4)$, on lit dans la table fournie en annexe que :

$$F_X(2) = 0{,}2381 \qquad F_X(4) = 0{,}6288$$

On en déduit que :

$$\Pr(2{,}57 < X \leq 4) = 0{,}6288 - 0{,}2381 = 0{,}3907$$

6 Loi normale

1. La loi normale est une loi continue et donc la probabilité associée à une réalisation particulière est nulle $\Pr(X = 2{,}57) = 0$.

$$\begin{aligned}\Pr(|X| \geq 1) &= 1 - \Pr(|X| \leq 1) \\ &= 1 - \Pr(-1 \leq X \leq 1) \\ &= 1 - \Pr(X \leq 1) + \Pr(X \leq -1)\end{aligned}$$

Si $X \sim \mathcal{N}(2,2)$ alors $(X-2)/\sqrt{2}$ suit une loi normale centrée réduite $\mathcal{N}(0,1)$. Par conséquent :

$$\begin{aligned}\Pr(|X| \geq 1) &= 1 - \Pr(X \leq 1) + \Pr(X \leq -1) \\ &= 1 - \Pr\left(\frac{X-2}{\sqrt{2}} \leq \frac{1-2}{\sqrt{2}}\right) \\ &\quad + \Pr\left(\frac{X-2}{\sqrt{2}} \leq \frac{-1-2}{\sqrt{2}}\right) \\ &= 1 - \Phi(-0{,}7071) + \Phi(-2{,}1213)\end{aligned}$$

où $\Phi(.)$ désigne la fonction de répartition de la loi normale centrée réduite. Par lecture de la table de la loi normale centrée réduite, il vient :

$$\begin{aligned}\Phi(-0{,}7071) &= 1 - \Phi(0{,}7071) \simeq 1 - \Phi(0{,}71) \\ &\simeq 1 - 0{,}761148 = 0{,}238852\end{aligned}$$

La valeur précise, obtenue avec un logiciel (Matlab), est égale à $\Phi(-0{,}7071) = 0{,}2398$. De la même façon, par lecture de la table de la loi normale centrée réduite, il vient :

$$\begin{aligned}\Phi(-2{,}1213) &= 1 - \Phi(2{,}1213) \simeq 1 - \Phi(2{,}12) \\ &\simeq 1 - 0{,}982997 = 0{,}017003\end{aligned}$$

La valeur précise obtenue avec un logiciel est égale à $\Phi(-2{,}1213) = 0{,}0169$. Par conséquent :

$$\Pr(|X| \geq 1) \simeq 1 - 0{,}238852 + 0{,}017003 = 0{,}778151$$

Avec un logiciel, on obtient $\Pr(|X| \geq 1) = 0{,}7772$.

2. Soit $F_X^{-1}(\alpha)$ le quantile d'ordre α de la loi normale $\mathcal{N}(2,2)$ tel que :

$$\Pr\left(X \leq F_X^{-1}(\alpha)\right) = F_X\left(F_X^{-1}(\alpha)\right) = \alpha$$

Étant données les propriétés de la fonction de répartition de la loi normale, on sait que $F_X(-\infty) = 0$. Donc le quantile d'ordre $\alpha = 0$ de la loi normale $\mathcal{N}(2,2)$ est égal à $-\infty$, *i.e.* $F_X^{-1}(0) = -\infty$. La définition du quantile à $\alpha = 95$ % est la suivante :

$$\Pr\left(X \leq F_X^{-1}(0{,}95)\right) = 0{,}95$$

Si $X \sim \mathcal{N}(2,2)$ alors $(X-2)/\sqrt{2}$ suit une loi normale $\mathcal{N}(0,1)$. Dès lors :

$$\Pr\left(X \leq F_X^{-1}(0{,}95)\right) = \Pr\left(\frac{X-2}{\sqrt{2}} \leq \frac{F_X^{-1}(0{,}95)-2}{\sqrt{2}}\right) = \Phi\left(F_X^{-1}(0{,}95)\right) = 0{,}95$$

On a donc une relation entre le quantile à 95 % de la loi normale $\mathcal{N}(2,2)$ et le quantile à 95 % de la loi normale centrée réduite :

$$F_X^{-1}(0{,}95) = 2 + \sqrt{2} \times \Phi^{-1}(0{,}95)$$

Par lecture de la table de la loi normale centrée réduite, il vient $\Phi^{-1}(0{,}95) \simeq 1{,}65$. Avec un logiciel on obtient $\Phi^{-1}(0{,}95) = 1{,}6449$. Dès lors :

$$F_X^{-1}(0{,}95) \simeq 2 + \sqrt{2} \times 1{,}65 = 4{,}333$$

ou de façon plus précise avec un logiciel $F_X^{-1}(0{,}95) = 4{,}3262$.

3. La définition du quantile à $\alpha = 1$ % de la loi normale $\mathcal{N}(2,2)$ est la suivante :

$$\Pr\left(X \leq F_X^{-1}(0{,}01)\right) = 0{,}01$$

Si $X \sim \mathcal{N}(2,2)$ alors $(X-2)/\sqrt{2}$ suit une loi normale centrée réduite $\mathcal{N}(0,1)$. On a donc une relation entre le quantile à 1 % de la loi normale $\mathcal{N}(2,2)$ et le quantile à 1 % de la loi normale centrée réduite :

$$F_X^{-1}(0{,}01) = 2 + \sqrt{2} \times \Phi^{-1}(0{,}01) = 2 - \sqrt{2} \times \Phi^{-1}(0{,}99)$$

Par lecture de la table de la loi normale centrée réduite, il vient $\Phi^{-1}(0{,}99) \simeq 2{,}33$. Avec un logiciel on obtient $\Phi^{-1}(0{,}99) = 2{,}3263$. Dès lors :

$$F_X^{-1}(0{,}01) \simeq 2 - \sqrt{2} \times 2{,}33 = -1{,}2951$$

ou de façon plus précise avec un logiciel $F_X^{-1}(0{,}01) = -1{,}2900$.

7 Loi Binomiale

1. La variable X_n suit une loi Binomiale $\mathcal{B}(n,p)$ avec $p = 2/8 = 1/4$, donc

$$\mathbb{E}(X_n) = np = \frac{n}{4}$$

$$\mathbb{V}(X_n) = np(1-p) = \frac{n}{4}\left(1-\frac{1}{4}\right) = \frac{3n}{16}$$

On en déduit que :

$$\mathbb{E}(F_n) = \mathbb{E}\left(\frac{X_n}{n}\right) = \frac{1}{n}\mathbb{E}(X_n) = \frac{1}{4}$$

$$\mathbb{V}(F_n) = \mathbb{V}\left(\frac{X_n}{n}\right) = \frac{1}{n^2}\mathbb{V}(X_n) = \frac{3}{16n}$$

2. On suppose que $n = 10\,000$. Dès lors, il vient :

$$\mathbb{E}(F_n) = \frac{1}{4} \qquad \mathbb{V}(F_n) = \frac{3}{160\,000}$$

On en déduit que :

$$\Pr(F_n \in\,]0{,}22; 0{,}26[) = \Pr(F_n - \mathbb{E}(F_n) \in\,]-0{,}03; 0{,}01[) > \Pr(|F_n - \mathbb{E}(F_n)| < 0{,}01)$$

ou encore :

$$\Pr(F_n \in\,]0{,}22; 0{,}26[) > 1 - \Pr(|F_n - \mathbb{E}(F_n)| \geq 0{,}01)$$

D'après l'inégalité de Tchebychev (▶ chapitre 8), nous avons :

$$\Pr(|F_n - \mathbb{E}(F_n)| \geq 0{,}01) \leq \frac{\mathbb{V}(F_n)}{0{,}01^2}$$

On obtient :

$$\Pr(F_n \in\,]0{,}22; 0{,}26[) \geq 1 - \frac{\mathbb{V}(F_n)}{0{,}01^2} = 1 - \frac{3}{160\,000 \times 0{,}01^2}$$

ou encore :

$$\Pr(F_n \in\,]0{,}22; 0{,}26[) \geq \frac{13}{16}$$

3. Le nombre minimal n de tirages nécessaires pour que la probabilité de l'événement $F_n \in\,]0{,}22; 0{,}26[$ soit au moins égale à 0,99 est tel que :

$$\Pr(F_n \in\,]0{,}22; 0{,}26[) \geq 1 - \frac{\mathbb{V}(F_n)}{0{,}01^2} = 0{,}99$$

ou encore :

$$1 - \frac{3}{0{,}01^2 \times 16n} = 0{,}99$$

On en déduit la valeur minimale de n :

$$n = \frac{3}{16 \times 0{,}01^3} = 187\,500$$

8 Loi normale

Soit X une variable aléatoire dont la loi de probabilité est la loi normale de moyenne 12 et d'écart-type 4.

$$X \sim \mathcal{N}(12,16) \Leftrightarrow \frac{X-12}{4} \sim \mathcal{N}(0,1)$$

On note $\Phi(.)$ la fonction de répartition de la loi normale centrée réduite.

1. Les probabilités de ces événements sont respectivement égales à :

$$\Pr(X = 2) = 0$$

puisque, pour une variable aléatoire continue, la probabilité d'être en un point est nulle.

$$\Pr(X < 16) = \Pr\left(\frac{X-12}{4} < \frac{16-12}{4}\right) = \Phi(1) = 0{,}8413$$

$$\begin{aligned}\Pr(X > 20) &= 1 - \Pr(X \leq 20)\\ &= 1 - \Pr\left(\frac{X-12}{4} < \frac{20-12}{4}\right)\\ &= 1 - \Phi(2)\\ &= 0{,}0228\end{aligned}$$

$$\Pr(X < 0) = \Pr\left(\frac{X-12}{4} < \frac{0-12}{4}\right) = \Phi(-3) = 0{,}0013$$

2. On cherche le nombre e tel que la probabilité de réalisation de l'événement $(|X-12|>e)$ soit égale à 0,01.

$$\begin{aligned}\Pr(|X-12|>e) &= 1-\Pr(|X-12|\le e)\\ &= 1-\Pr(-e\le X-12\le e)\\ &= 1-\Pr\left(-\frac{e}{4}\le\frac{X-12}{4}\le\frac{e}{4}\right)\\ &= 1-\Pr\left(\frac{X-12}{4}\le\frac{e}{4}\right)\\ &\quad+\Pr\left(\frac{X-12}{4}\le-\frac{e}{4}\right)\\ &= 1-\Phi\left(\frac{e}{4}\right)+\Phi\left(-\frac{e}{4}\right)\end{aligned}$$

Puisque $\Phi(-x)=1-\Phi(x)$, il vient

$$\Pr(|X-12|>e)=2-2\Phi\left(\frac{e}{4}\right)=0{,}01$$

On en déduit que :

$$\Phi\left(\frac{e}{4}\right)=\frac{2-0{,}01}{2}\iff e=4\times\Phi^{-1}\left(\frac{2-0{,}01}{2}\right)$$

La constante e est égale à :

$$e=4\times\Phi^{-1}(0{,}9950)=4\times2{,}5758=10{,}3033$$

3. La variable Y suit une loi normale avec :

$$\mathbb{E}(Y)=\mathbb{E}(aX+b)=a\mathbb{E}(X)+b=12a+b$$
$$\mathbb{V}(Y)=\mathbb{V}(aX+b)=a^2\mathbb{V}(X)=16a^2$$

Ainsi, on obtient :

$$Y\sim\mathcal{N}\left(12a+b,16a^2\right)\Leftrightarrow\frac{Y-12a-b}{4a}\sim\mathcal{N}(0,1)$$

Nous savons que :

$$\Pr(Y<24)=\Phi\left(\frac{24-12a-b}{4a}\right)=0{,}2266$$

$$\begin{aligned}\Pr(Y>42) &= 1-\Pr(Y\le42)\\ &= 1-\Phi\left(\frac{42-12a-b}{4a}\right)=0{,}0068\end{aligned}$$

On obtient un système à deux équations et deux inconnues :

$$\frac{24-12a-b}{4a}=\Phi^{-1}(0{,}2266)\simeq-0{,}75$$
$$\frac{42-12a-b}{4a}=\Phi^{-1}(0{,}9320)\simeq1{,}5$$

On en déduit que :

$$a=2,\quad b=6$$

9 Loi binomiale et loi de Poisson

On considère les nombres x_0 et x_1 définis comme suit :

– x_0 est la plus grande valeur entière de x telle que :

$$\Pr(X\le x)\le0{,}05$$

– x_1 est la plus petite valeur entière de x telle que :

$$\Pr(X\ge x)\le0{,}05$$

1. On suppose que la variable X suit une loi binomiale $\mathcal{B}(40;0{,}08)$. Déterminons x_0, la plus grande valeur entière de x telle que :

$$\Pr(X\le x)\le0{,}05$$

Nous savons que :

$$\Pr(X\le0)=0{,}0356$$
$$\Pr(X\le1)=0{,}1594$$

Donc $x_0=0$. De la même façon, déterminons x_1, la plus petite valeur entière de x telle que :

$$\Pr(X\ge x)\le0{,}05\iff1-\Pr(X<x)\le0{,}05$$

Ce qui peut se réécrire sous la forme :

$$\Pr(X\le x-1)\ge0{,}95$$

puisque X est définie sur le support $X(\Omega)=\mathbb{N}$. Nous savons que :

$$\Pr(X\le5)=0{,}9033$$
$$\Pr(X\le6)=0{,}9624$$

Donc $x-1\ge6$ ou $x\ge7$. La plus petite valeur entière de x telle que $\Pr(X\ge x)\le0{,}05$ est égale à $x_1=7$.

2. On suppose que la variable X suit une loi de Poisson $\mathcal{P}(12)$. Déterminons x_0, la plus grande valeur entière de x telle que :

$$\Pr(X\le x)\le0{,}05$$

Nous savons que :

$$\Pr(X\le6)=0{,}0458$$
$$\Pr(X\le7)=0{,}0895$$

Donc $x_0=6$. De la même façon, déterminons x_1, la plus petite valeur entière de x telle que :

$$\Pr(X\ge x)\le0{,}05\iff1-\Pr(X<x)\le0{,}05$$

Ce qui peut se réécrire sous la forme :

$$\Pr(X\le x-1)\ge0{,}95$$

puisque X est définie sur le support $X(\Omega)=\mathbb{N}$. Nous savons que :

$$\Pr(X\le17)=0{,}9370$$
$$\Pr(X\le18)=0{,}9626$$

Donc $x-1\ge18$ ou $x\ge19$. La plus petite valeur entière de x telle que $\Pr(X\ge x)\le0{,}05$ est égale à $x_1=19$.

10 Loi de Fisher et loi du khi-deux

On considère les nombres x_0 et x_1 définis comme suit :

$$\Pr(X<x_0)=0{,}05$$
$$\Pr(X>x_1)=0{,}05$$

1. On suppose que la variable X suit une de Fisher $\mathcal{F}(4{,}20)$. Soit $F_X(x;n,m)$ la fonction de répartition de la loi $\mathcal{F}(n,m)$. En utilisant un logiciel de statistique, on obtient directement :

$$\Pr(X<x_0)=0{,}05\iff x_0=F_X^{-1}(0{,}05;4{,}20)=0{,}1723$$
$$\Pr(X>x_1)=0{,}05\iff x_1=F_X^{-1}(0{,}95;4{,}20)=2{,}8661$$

2. On suppose que la variable X suit une loi du khi-deux $\chi^2(8)$. Soit $F_X(x;\nu)$ la fonction de répartition de la loi $\chi^2(\nu)$. En utilisant un logiciel de statistique, on obtient directement :

$$\Pr(X < x_0) = 0{,}05 \iff x_0 = F_X^{-1}(0{,}05;8) = 2{,}7326$$

$$\Pr(X > x_1) = 0{,}05 \iff x_1 = F_X^{-1}(0{,}95;8) = 15{,}5073$$

Chapitre 8

1 Convergence

a. Faux. La convergence en probabilité implique la convergence en loi, mais la réciproque n'est pas toujours vraie.

b. Vrai. La convergence en probabilité implique la convergence en loi au sens où si $Y_n \xrightarrow{p} Y \iff Y_n - Y \xrightarrow{p} 0$, alors $Y_n \xrightarrow{d} Y$.

c. Vrai. La convergence presque sûre implique la convergence en probabilité et donc la convergence en loi.

d. Vrai. La convergence presque sûre implique la convergence en probabilité, mais la réciproque n'est pas toujours vraie.

e. Faux.

2

Si $Y_1,\ldots,Y_n$ sont des variables indépendantes et identiquement distribuées, alors :

a. Faux. Sans hypothèses supplémentaires sur les variables Y_i, on ne peut pas invoquer la loi forte des grands nombres.

b. Vrai. D'après la loi faible des grands nombres, la moyenne empirique converge en probabilité vers l'espérance.

c. Faux. Cette phrase ne veut rien dire.

d. Faux. La moyenne empirique converge en probabilité vers l'espérance, et donc sa distribution asymptotique est dégénérée (au sens strict).

e. Vrai. D'après le théorème central limite, $\sqrt{n}\left(\overline{Y}_n - \mathbb{E}(Y_i)\right)$ converge en loi vers une loi normale.

3

Si une variable Z_n est asymptotiquement normalement distribuée, cela signifie que :

a. Faux. Rien ne permet de dire que la loi asymptotique est $\mathcal{N}(0,1)$.

b. Vrai. Au sens strict, cela signifie que $\sqrt{n}(Z_n - \mathbb{E}(Z_n)) \xrightarrow{d} \mathcal{N}\left(m,\sigma^2\right)$.

c. Vrai. Lorsque $\sqrt{n}(Z_n - \mathbb{E}(Z_n)) \xrightarrow{d} \mathcal{N}\left(0,\sigma^2\right)$, on dit que Z_n est asymptotiquement normalement distribuée.

d. Faux. Rien ne permet de dire que la loi asymptotique est $\mathcal{N}(0,1)$.

e. Vrai. Si la variable Z_n est asymptotiquement normalement distribuée :

$$\sqrt{n}(Z_n - m) \xrightarrow{d} \mathcal{N}\left(0,\sigma^2\right)$$

Pour une dimension n grande et finie, on peut écrire que :

$$Z_n \overset{asy}{\approx} \mathcal{N}\left(m,\frac{\sigma^2}{n}\right)$$

4

Soit $Y_1,\ldots,Y_n$ une suite de n variables aléatoires et soit $\overline{Y}_n$ la moyenne empirique :

a. Vrai. C'est le théorème de Lindeberg-Levy qui s'applique dans ce cas.

b. Vrai. C'est le théorème de Lyapounov qui s'applique dans ce cas.

c. Vrai. Si les espérances sont identiques, c'est le théorème de Lindeberg–Feller qui s'applique. Si les espérances sont différentes, on applique le théorème de Lyapounov.

d. Vrai. C'est le théorème de Lyapounov qui s'applique dans ce cas.

e. Faux. Sans hypothèse particulière sur la forme de la dépendance (dépendance faible), le théorème central limite ne s'applique pas.

5 Convergences

1. Les variables X_1/σ et X_2/σ sont indépendantes et suivent toutes deux une loi normale centrée réduite. Dès lors, les variables X_1^2/σ^2 et X_2^2/σ^2 sont elles aussi indépendantes et suivent une loi du khi-deux à un degré de liberté (▸ chapitre 6). La variable Y^2/σ^2 est une somme de deux khi-deux indépendants, elle suit une distribution du khi-deux à 2 degrés de liberté :

$$\frac{Y^2}{\sigma^2} = \frac{X_1^2}{\sigma^2} + \frac{X_2^2}{\sigma^2} \sim \chi^2(2)$$

D'après les propriétés de la loi du khi-deux, on sait que si $Z \sim \chi^2(v)$ alors $\mathbb{E}(Z) = v$ et $\mathbb{V}(Z) = 2v$. On en déduit que :

$$\mathbb{E}\left(\frac{Y^2}{\sigma^2}\right) = \frac{1}{\sigma^2}\mathbb{E}\left(Y^2\right) = 2 \iff \mathbb{E}\left(Y^2\right) = 2\sigma^2$$

$$\mathbb{V}\left(\frac{Y^2}{\sigma^2}\right) = \frac{1}{\sigma^4}\mathbb{V}\left(Y^2\right) = 4 \iff \mathbb{V}\left(Y^2\right) = 4\sigma^4$$

2. Calculons $\mathbb{E}\left(\widehat{\sigma}^2\right)$ et $\mathbb{V}\left(\widehat{\sigma}^2\right)$. Sous l'hypothèse d'un échantillon i.i.d., il vient :

$$\mathbb{E}\left(\widehat{\sigma}^2\right) = \frac{1}{2n}\sum_{i=1}^{n}\mathbb{E}\left(Y_i^2\right) = \frac{n \times 2\sigma^2}{2n} = \sigma^2$$

$$\mathbb{V}\left(\widehat{\sigma}^2\right) = \mathbb{V}\left(\frac{1}{2n}\sum_{i=1}^{n}Y_i^2\right) = \frac{1}{4n^2}\sum_{i=1}^{n}\mathbb{V}\left(Y_i^2\right) = \frac{n \times 4\sigma^4}{4n^2} = \frac{\sigma^4}{n}$$

Dès lors :

$$\mathbb{E}\left(\widehat{\sigma}^2\right) = \sigma^2$$

$$\lim_{n\to\infty}\mathbb{V}\left(\widehat{\sigma}^2\right) = \lim_{n\to\infty}\frac{\sigma^4}{n} = 0$$

On en déduit que $\widehat{\sigma}^2$ converge vers σ^2 :

$$\widehat{\sigma}^2 \xrightarrow{p} \sigma^2$$

3. Les variables $Y_1^2,\ldots,Y_n^2$ sont i.i.d. avec $\mathbb{E}\left(Y_i^2\right) = 2\sigma^2$ et $\mathbb{V}\left(Y_i^2\right) = 4\sigma^4$. Par application du théorème central limite, il vient :

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}Y_i^2 - 2\sigma^2\right) \xrightarrow{d} \mathcal{N}\left(0,4\sigma^4\right)$$

Or, l'estimateur $\widehat{\sigma}^2$ est défini par :

$$\widehat{\sigma}^2 = \frac{1}{2n}\sum_{i=1}^{n}Y_i^2$$

En utilisant la méthode delta pour une fonction $g(x) = x/2$, il vient :

$$\sqrt{n}\left(\widehat{\sigma}^2 - \sigma^2\right) \xrightarrow{d} \mathcal{N}\left(0,\sigma^4\right)$$

6 Théorème central limite

1. Soit X_i la variable binaire valant 1 si l'abonné i est connecté et 0 sinon. La variable X_i suit une distribution de Bernouilli de probabilité $p = \Pr(X_i = 1) = 0{,}20$. Les variables X_i sont indépendantes, donc la variable X définie par :

$$X = \sum_{i=1}^{n} X_i$$

avec $n = 5\,000$ suit une loi binomiale $\mathcal{B}(5\,000; 0{,}20)$. Par définition :

$$\mathbb{E}(X) = np = 5\,000 \times 0{,}2 = 1000$$
$$\mathbb{V}(X) = np(1-p) = 5\,000 \times 0{,}2 \times 0{,}8 = 800$$

2. Les variables X_i sont i.i.d. avec $\mathbb{E}(X_i) = p = 0{,}2$ et $\mathbb{V}(X_i) = p(1-p) = 0{,}16$, par application du théorème central limite de Lindeberg-Levy il vient :

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - p\right) \xrightarrow{d} \mathcal{N}(0, p(1-p))$$

Pour n grand mais fini, on a donc :

$$\sqrt{n}\left(\frac{X}{n} - p\right) \overset{asy}{\approx} \mathcal{N}(0, p(1-p))$$

ou encore :

$$\frac{X}{n} \overset{asy}{\approx} \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

Finalement, il vient :

$$X \overset{asy}{\approx} \mathcal{N}(np, np(1-p)) \Leftrightarrow \frac{X - np}{\sqrt{np(1-p)}} \overset{asy}{\approx} \mathcal{N}(0,1)$$

Si l'on admet que $n = 5\,000$ est une taille suffisamment importante pour que cette approximation soit valide, la variable Y suit asymptotiquement une loi normale centrée réduite :

$$Y = \frac{X - 1\,000}{\sqrt{800}} \overset{asy}{\approx} \mathcal{N}(0,1)$$

3. On cherche un seuil N tel que :

$$\Pr(X \geq N) \leq 0{,}975$$

On sait que :

$$\Pr(X \geq N) = \Pr\left(Y \geq \frac{N - 1\,000}{\sqrt{800}}\right) \leq 0{,}025$$

Si l'on admet que Y suit approximativement une loi normale centrée réduite, il vient :

$$1 - \Pr\left(Y < \frac{N - 1\,000}{\sqrt{800}}\right) \leq 0{,}025$$
$$\Longleftrightarrow \Phi\left(\frac{N - 1\,000}{\sqrt{800}}\right) \geq 0{,}975$$

où $\Phi(.)$ désigne la fonction de répartition de la loi normale centrée réduite. Nous savons que :

$$0{,}975 = \Phi\left(\Phi^{-1}(0{,}975)\right) = \Phi(1{,}96)$$

Ainsi, on obtient une relation du type :

$$\Phi\left(\frac{N - 1\,000}{\sqrt{800}}\right) \geq \Phi(1{,}96)$$

Puisque la fonction de répartition $\Phi(.)$ est croissante :

$$\frac{N - 1\,000}{\sqrt{800}} \geq 1{,}96$$

On en déduit :

$$N \geq 1{,}96 \times \sqrt{800} + 1\,000 = 1\,054$$

7 Convergences

Soit (X_n) une suite de variables aléatoires discrètes telles que $\forall n \geq 2$:

$$\Pr(X_n = -n) = \Pr(X_n = n) = \frac{1}{2n^2}$$
$$\Pr(X_n = 0) = 1 - \frac{1}{n^2}$$

1. On fait l'hypothèse que la limite en probabilité de la suite (X_n) est égale à 0.

$$X_n \xrightarrow{p} 0$$

Cette hypothèse est vraie si et seulement si $\forall \varepsilon$:

$$\lim_{n\to\infty} \Pr(|X_n - 0| > \varepsilon) = 0$$

ou de façon équivalente lorsque :

$$\lim_{n\to\infty} \Pr(|X_n - 0| < \varepsilon) = 1$$

Or, par définition, si l'on fait tendre ε vers 0, il vient :

$$\Pr(|X_n - 0| < \varepsilon) = \Pr(|X_n| < \varepsilon) = \Pr(X_n = 0) = 1 - \frac{1}{n^2}$$

Ainsi, on obtient :

$$\lim_{n\to\infty} \Pr(|X_n - 0| < \varepsilon) = \lim_{n\to\infty} 1 - \frac{1}{n^2} = 1$$

On en déduit que :

$$X_n \xrightarrow{p} 0$$

2. La suite (X_n) converge en moyenne quadratique vers 0 si et seulement si :

$$\lim_{n\to\infty} \mathbb{E}(|X_n - 0|^2) = \lim_{n\to\infty} \mathbb{E}(|X_n|^2) = \lim_{n\to\infty} \mathbb{E}(X_n^2) = 0$$

Or, par définition :

$$\mathbb{E}(X_n^2) = \sum i^2 \Pr(X = i)$$
$$= 0^2 \Pr(X = 0) + \sum_{i>0} i^2 \Pr(X = i) + \sum_{i<0} i^2 \Pr(X = i)$$
$$= \sum_{i>0} i^2 \Pr(X = i) + \sum_{i>0} i^2 \Pr(X = -i)$$
$$= \sum_{i>0} i^2 \frac{2}{2i^2} > 0$$

Donc il n'y a pas de convergence en moyenne quadratique vers 0.

8 Théorème central limite

1. Soit X_i la variable binaire valant 1 si le passager i se présente à l'embarquement et 0 sinon. La variable X_i suit une distribution de Bernouilli de probabilité $p = \Pr(X_i = 1) = 0{,}90$. Les variables X_i sont indépendantes, donc la variable S_n définie par :

$$S_n = \sum_{i=1}^{n} X_i$$

suit une loi binomiale $\mathcal{B}(n; 0{,}90)$. Par définition :

$$\mathbb{E}(S_n) = np = 0{,}9n$$

$$\mathbb{V}(S_n) = np(1-p) = 0{,}09n$$

2. Les variables X_i sont i.i.d. avec $\mathbb{E}(X_i) = p = 0{,}9$ et $\mathbb{V}(X_i) = p(1-p) = 0{,}09$. Par application du théorème central limite de Lindeberg-Levy il vient :

$$\sqrt{n}\left(\frac{S_n}{n} - p\right) \overset{d}{\to} \mathcal{N}(0, p(1-p))$$

Pour n grand mais fini, on a donc :

$$\sqrt{n}\left(\frac{S_n}{n} - p\right) \overset{asy}{\approx} \mathcal{N}(0, p(1-p))$$

ou encore :

$$\frac{S_n}{n} \overset{asy}{\approx} \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

Finalement, il vient :

$$S_n \overset{asy}{\approx} \mathcal{N}(np, np(1-p)) \Leftrightarrow \frac{S_n - np}{\sqrt{np(1-p)}} \overset{asy}{\approx} \mathcal{N}(0,1)$$

ou encore

$$\frac{S_n - 0{,}9n}{0{,}3\sqrt{n}} \overset{asy}{\approx} \mathcal{N}(0,1)$$

Le directeur commercial de la compagnie aimerait connaitre la valeur maximale de n telle que :

$$\Pr(S_n \leq 300) \geq 0{,}99$$

Par définition :

$$\Pr(S_n \leq 300) = \Pr\left(\frac{S_n - 0{,}9n}{0{,}3\sqrt{n}} \leq \frac{300 - 0{,}9n}{0{,}3\sqrt{n}}\right) = \Phi\left(\frac{300 - 0{,}9n}{0{,}3\sqrt{n}}\right)$$

où $\Phi(.)$ désigne la fonction de répartition de la loi normale centrée réduite. On a donc :

$$\Phi\left(\frac{300 - 0{,}9n}{0{,}3\sqrt{n}}\right) \geq \Phi\left(\Phi^{-1}(0{,}99)\right) = \Phi(2{,}3263)$$

Puisque la fonction $\Phi(.)$ est croissante :

$$\frac{300 - 0{,}9n}{0{,}3\sqrt{n}} \geq 2{,}3263$$

On obtient l'équation :

$$-0{,}9n - 0{,}6979\sqrt{n} + 300 \geq 0$$

En posant $x = \sqrt{n}$, on obtient une fonction du second degré à étudier et l'on cherche le plus grand x pour lequel elle est positive. Le plus grand n pour lequel cette quantité est positive est 316.

Chapitre 9

1 Propriétés d'un estimateur

a. Faux. Un estimateur sans biais n'est pas nécessairement efficace.

b. Faux. Un estimateur sans biais n'est pas nécessairement convergent. Il n'est convergent que si variance tend vers zéro lorsque la taille d'échantillon tend vers l'infini.

c. Vrai. Puisque la convergence presque sûre implique la convergence en probabilité.

d. Faux. Un estimateur biaisé peut être convergent s'il est asymptotiquement non biaisé.

e. Vrai. La comparaison de la variance de l'estimateur à la borne FDCR ne s'applique que pour des estimateurs sans biais. Un estimateur efficace est donc sans biais.

2 Variance empirique

a. Vrai. La variance empirique (corrigée ou non corrigée) est un estimateur de la variance.

b. Vrai.

c. Faux. Seule une transformée de la variance empirique corrigée a une distribution exacte du khi-deux.

d. Faux. Le nombre de degrés de liberté est $n-1$ et non n.

e. Faux. En raison de la correction de petit échantillon, on a $S_n^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$.

3 Comparaison d'estimateurs

a. Vrai. L'estimateur $\widehat{\theta}_1$ est préféré à $\widehat{\theta}_2$ si sa variance est plus faible.

b. Faux. Ce n'est parce que $\mathbb{V}(\widehat{\theta}_1) < \mathbb{V}(\widehat{\theta}_2)$, que la variance de $\widehat{\theta}_1$ atteint la borne FDCR.

c. Vrai. Si la variance de $\widehat{\theta}_1$ atteint la borne FDCR, alors la variance de tout estimateur sans biais est nécessairement égale ou supérieure à celle de l'estimateur $\widehat{\theta}_1$.

d. Vrai. L'optimalité implique généralement l'efficacité, mais la réciproque n'est pas vraie puisque l'efficacité requiert d'effectuer un certain nombre d'hypothèses sur la distribution des variables de l'échantillon.

e. Faux. La convergence au sens faible n'implique pas nécessairement l'efficacité.

4 Intervalle de confiance

a. Vrai. C'est la définition d'un intervalle de confiance.

b. Vrai.

c. Vrai. En effet, plus l'estimateur est précis, plus la réalisation de l'intervalle de confiance sera concentrée autour de la vraie valeur du paramètre.

d. Vrai. En général, le niveau de risque α est inférieur à 50 %.

e. Faux. C'est la réalisation d'un intervalle de confiance pour un échantillon donné qui se ramène à un segment de deux valeurs réelles.

5 Estimation et loi de Rayleigh

1. La variable Y^2/σ^2 peut s'écrire sous la forme :

$$\left(\frac{Y}{\sigma}\right)^2 = \left(\frac{X_1}{\sigma}\right)^2 + \left(\frac{X_2}{\sigma}\right)^2$$

Les variables X_1/σ et X_2/σ suivent des lois normales centrées réduites $\mathcal{N}(0,1)$. Par conséquent, les variables X_1^2/σ^2 et X_2^2/σ^2 ont une distribution du khi-deux à un degré de liberté. Puisque ces variables sont indépendantes, la somme $X_1^2/\sigma^2 + X_2^2/\sigma^2$ admet une distribution du khi-deux à deux degrés de liberté.

$$\left(\frac{Y}{\sigma}\right)^2 \sim \chi^2(2)$$

Etant données les propriétés de la loi du khi-deux, on sait que si $X \sim \chi^2(v)$, alors $\mathbb{E}(X) = v$ et $\mathbb{V}(X) = 2v$. On en déduit que :

$$\mathbb{E}\left(\frac{Y^2}{\sigma^2}\right) = \frac{1}{\sigma^2}\mathbb{E}\left(Y^2\right) = 2 \iff \mathbb{E}\left(Y^2\right) = 2\sigma^2$$

$$\mathbb{V}\left(\frac{Y^2}{\sigma^2}\right) = \frac{1}{\sigma^4}\mathbb{V}\left(Y^2\right) = 4 \iff \mathbb{V}\left(Y^2\right) = 4\sigma^4$$

2. On étudie la quantité $\mathbb{E}(\widehat{\sigma}^2)$:

$$\mathbb{E}\left(\widehat{\sigma}^2\right) = \mathbb{E}\left(\frac{1}{2n}\sum_{i=1}^n Y_i^2\right) = \frac{1}{2n}\sum_{i=1}^n \mathbb{E}\left(Y_i^2\right) = \frac{n \times 2\sigma^2}{2n} = \sigma^2$$

puisque les variables $Y_1,\ldots,Y_n$ sont i.i.d., de même loi que Y, avec $\mathbb{E}\left(Y_i^2\right) = 2\sigma^2$. L'estimateur $\widehat{\sigma}^2$ est sans biais.

3. Nous savons que l'estimateur est sans biais :

$$\mathbb{E}\left(\widehat{\sigma}^2\right) = \sigma^2$$

Par ailleurs :

$$\mathbb{V}\left(\widehat{\sigma}^2\right) = \mathbb{V}\left(\frac{1}{2n}\sum_{i=1}^n Y_i^2\right) = \frac{1}{4n^2}\sum_{i=1}^n \mathbb{V}\left(Y_i^2\right) = \frac{n \times 4\sigma^4}{4n^2} = \frac{\sigma^4}{n}$$

puisque les variables $Y_1,\ldots,Y_n$ sont i.i.d. avec $\mathbb{V}\left(Y_i^2\right) = 4\sigma^4$. Dès lors :

$$\lim_{n\to\infty}\mathbb{V}\left(\widehat{\sigma}^2\right) = \lim_{n\to\infty}\frac{\sigma^4}{n} = 0$$

On peut en déduire que l'estimateur est convergent (au sens faible) :

$$\widehat{\sigma}^2 \xrightarrow{p} \sigma^2$$

4. Les variables $Y_1^2,\ldots,Y_n^2$ sont i.i.d. avec $\mathbb{E}\left(Y_i^2\right) = 2\sigma^2$ et $\mathbb{V}\left(Y_i^2\right) = 4\sigma^4$. Par application du théorème central limite de Lindeberg-Levy, il vient :

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n Y_i^2 - 2\sigma^2\right) \xrightarrow{d} \mathcal{N}\left(0,4\sigma^4\right)$$

Or, l'estimateur $\widehat{\sigma}^2$ vérifie :

$$\widehat{\sigma}^2 = \frac{1}{2n}\sum_{i=1}^n Y_i^2 = g\left(\frac{1}{n}\sum_{i=1}^n Y_i^2\right)$$

où la fonction $g(.)$ est définie par $g(x) = x/2$. Cette fonction est continue et continument différentiable, avec $\partial g(x)/\partial x = 1/2$. Par application de la méthode delta (▸ chapitre 6), il vient :

$$\sqrt{n}\left(g\left(\frac{1}{n}\sum_{i=1}^n Y_i^2\right) - g\left(2\sigma^2\right)\right) \xrightarrow{d} \mathcal{N}\left(0,\left(\left.\frac{\partial g(z)}{\partial z}\right|_{2\sigma^2}\right) \times 4\sigma^4\right)$$

ou de façon équivalente :

$$\sqrt{n}\left(\widehat{\sigma}^2 - \sigma^2\right) \xrightarrow{d} \mathcal{N}\left(0,\left(\frac{1}{2}\right)^2 \times 4\sigma^4\right)$$

Après simplification, on obtient :

$$\sqrt{n}\left(\widehat{\sigma}^2 - \sigma^2\right) \xrightarrow{d} \mathcal{N}\left(0,\sigma^4\right)$$

5. Le résultat précédent signifie que pour une taille d'échantillon n suffisament grande, mais finie :

$$\widehat{\sigma}^2 \overset{asy}{\approx} \mathcal{N}\left(\sigma^2,\frac{\sigma^4}{n}\right)$$

La variance asymptotique de l'estimateur $\widehat{\sigma}^2$ est donc égale à :

$$\mathbb{V}_{asy}\left(\widehat{\sigma}^2\right) = \frac{\sigma^4}{n}$$

6. Ici nous avons :

$$\mathbb{V}\left(\widehat{\sigma}^2\right) = \frac{\sigma^4}{n} = \mathrm{I}_n^{-1}(\theta)$$

La variance de l'estimateur $\widehat{\sigma}^2$ atteint la borne FDCR : l'estimateur est efficace.

6 Loi exacte et loi asymptotique

1. Puisque $Z_i \sim \mathcal{N}(0,1)$, la variable Z_i^2 suit une loi du khi-deux à un degré de liberté. Les variables Z_i^2 sont indépendantes, donc la variable $D_n = \sum_{i=1}^n Z_i^2$ suit une loi du khi-deux à n degrés de liberté : $D_n \sim \chi^2(n)$.

2. Puisque $Z_i^2 \sim \chi^2(1)$, on a $\mathbb{E}\left(Z_i^2\right) = 1$ et $\mathbb{V}\left(Z_i^2\right) = 2$. Les variables Z_i^2 sont i.i.d., donc nous pouvons appliquer le thèorème central limite de Lindeberg-Levy :

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n Z_i^2 - 1\right) = \sqrt{n}\left(\frac{D_n}{n} - 1\right) \xrightarrow{d} \mathcal{N}(0,2)$$

3. On note $F_n(x)$ la fonction de répartition d'une loi $\chi^2(n)$. En utilisant la loi exacte $\chi^2(100)$, on obtient :

$$\begin{aligned}\Pr(D_n > 118{,}49) &= 1 - \Pr(D_n \leq 118{,}49)\\ &= 1 - F_{100}(118{,}49)\\ &= 1 - 0{,}90 = 0{,}1\end{aligned}$$

où $F_{100}(.)$ désigne la fonction de répartition d'une loi du khi-deux à 100 degrés de liberté. La loi asymptotique de D_n peut s'écrire sous la forme suivante :

$$\frac{D_n}{n} \overset{asy}{\approx} \mathcal{N}\left(1,\frac{2}{n}\right) \quad \text{ou} \quad D_n \overset{asy}{\approx} \mathcal{N}(n,2n)$$

Si l'on admet que pour $n = 100$ cette approximation est valide, il vient :

$$\begin{aligned}\Pr(D_n > 118{,}49) &= 1-\Pr(D_n \leq 118{,}49)\\ &= 1-\Pr\left(\frac{D_n - n}{\sqrt{2n}} \leq \frac{118{,}49 - n}{\sqrt{2n}}\right)\\ &= 1-\Phi\left(\frac{118{,}49-100}{\sqrt{200}}\right)\\ &= 1-\Phi(1{,}308)\\ &\simeq 0{,}0954\end{aligned}$$

où $\Phi(.)$ est la fonction de répartition de la loi normale centrée réduite.

7 Comparaison d'estimateurs

1. Calculons $\mathbb{E}\left(\widehat{\theta}_1\right)$ et $\mathbb{E}\left(\widehat{\theta}_2\right)$:

$$\mathbb{E}\left(\widehat{\theta}_1\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X) = \theta$$

$$\mathbb{E}\left(\widehat{\theta}_2\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(X_i^2\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(X^2\right)$$

Or on sait que

$$\mathbb{E}\left(X^2\right) = \mathbb{V}(X) + \mathbb{E}(X)^2 = \theta - \theta^2 + \theta^2 = \theta$$

Il vient donc :

$$\mathbb{E}\left(\widehat{\theta}_1\right) = \theta \quad \mathbb{E}\left(\widehat{\theta}_2\right) = \theta$$

Les deux estimateurs sont sans biais.

2. Calculons $\mathbb{V}\left(\widehat{\theta}_1\right)$ et $\mathbb{V}\left(\widehat{\theta}_2\right)$. Puisque les variables X_i sont i.i.d., on a :

$$\mathbb{V}\left(\widehat{\theta}_1\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X) = \frac{\theta-\theta^2}{n}$$

$$\mathbb{V}\left(\widehat{\theta}_2\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}\left(X_i^2\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}\left(X^2\right) = \frac{2\theta^2-2\theta^4}{n}$$

Dans les deux cas, on a :

$$\mathbb{E}\left(\widehat{\theta}_1\right) = \mathbb{E}\left(\widehat{\theta}_2\right) = \theta$$

$$\lim_{n\to\infty}\mathbb{V}\left(\widehat{\theta}_1\right) = \lim_{n\to\infty}\mathbb{V}\left(\widehat{\theta}_2\right) = 0$$

Les deux estimateurs sont donc convergents au sens de la convergence en probabilité.

3. On sait que

$$\mathbb{V}\left(\widehat{\theta}_1\right) = \frac{\theta-\theta^2}{n}$$

$$\mathbb{V}\left(\widehat{\theta}_2\right) = \frac{2\theta^2-2\theta^4}{n}$$

Dès lors :

$$\frac{\mathbb{V}\left(\widehat{\theta}_1\right)}{\mathbb{V}\left(\widehat{\theta}_2\right)} = \frac{\theta(1-\theta)}{2\theta^2(1-\theta^2)} = \frac{1}{2\theta(1+\theta)}$$

avec $\theta \in [0,1]$. Donc on ne peut pas déterminer si $\mathbb{V}\left(\widehat{\theta}_1\right) > \mathbb{V}\left(\widehat{\theta}_2\right)$ ou si $\mathbb{V}\left(\widehat{\theta}_1\right) < \mathbb{V}\left(\widehat{\theta}_2\right)$ puisque la position du ratio $\mathbb{V}\left(\widehat{\theta}_1\right)/\mathbb{V}\left(\widehat{\theta}_2\right)$ par rapport à 1 dépend de la vraie valeur de θ qui est inconnue.

4. D'après le théorème central limite, on a :

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}X_i - \mathbb{E}(X)\right) \xrightarrow{d} \mathcal{N}(0,\mathbb{V}(X))$$

Donc

$$\sqrt{n}\left(\widehat{\theta}_1 - \alpha\right) \xrightarrow{d} \mathcal{N}\left(0,\theta-\theta^2\right)$$

De la même façon :

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2 - \mathbb{E}\left(X^2\right)\right) \xrightarrow{d} \mathcal{N}\left(0,\mathbb{V}\left(X^2\right)\right)$$

Donc

$$\sqrt{n}\left(\widehat{\theta}_2 - \alpha\right) \xrightarrow{d} \mathcal{N}(0,2\theta(1+\theta))$$

8 Estimation

1. Calculons l'espérance de l'estimateur $\widehat{\theta}$:

$$\begin{aligned}\mathbb{E}\left(\widehat{\theta}\right) &= \mathbb{E}\left(\ln(c) - \frac{1}{n}\sum_{i=1}^{n}\ln(X_i)\right)\\ &= \ln(c) - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(\ln(X_i))\\ &= \ln(c) - \frac{1}{n}\sum_{i=1}^{n}(\ln(c)-\theta)\\ &= \theta\end{aligned}$$

L'estimateur $\widehat{\theta}$ est sans biais.

2. La séquence de variables aléatoires i.i.d. $\ln(X_1),\ldots,\ln(X_n)$ vérifie $\mathbb{E}(\ln(X_i)) = \ln(c) - \theta$. D'après la loi faible des grands nombres :

$$\frac{1}{n}\sum_{i=1}^{n}\ln(X_i) \xrightarrow{p} \mathbb{E}(\ln(X_i)) = \ln(c) - \theta$$

En utilisant le *continuous mappin theorem* pour une fonction $g(z) = \ln(c) - z$, on obtient :

$$g\left(\frac{1}{n}\sum_{i=1}^{n}\ln(X_i)\right) \xrightarrow{p} g(\ln(c)-\theta)$$

ou de façon équivalente :

$$\ln(c) - \frac{1}{n}\sum_{i=1}^{n}\ln(X_i) \xrightarrow{p} \ln(c) - \ln(c) + \theta$$

Par conséquent :

$$\widehat{\theta} \xrightarrow{p} \theta$$

L'estimateur $\widehat{\theta}$ est convergent (au sens faible).

Chapitre 10

1 Vraisemblance et log-vraisemblance

a. Vrai. La log-vraisemblance d'un échantillon correspond au logarithme de la vraisemblance de l'échantillon.

b. Faux. Pour une variable continue, la vraisemblance d'un échantillon correspond à la densité jointe des variables de l'échantillon.

c. Faux. La log-vraisemblance d'un échantillon est égale à la somme des vraisemblances individuelles associées à chaque observation de cet échantillon.

d. Faux. La log-vraisemblance d'un échantillon est une constante puisqu'elle dépend des paramètres (quantités déterministes) et de la réalisation de l'échantillon.

e. Vrai. La vraisemblance d'un échantillon dépend de deux arguments : le vecteur de paramètres et les données de l'échantillon.

2 Estimateur du maximum de vraisemblance

a. Faux. L'estimateur du maximum de vraisemblance, comme tout estimateur, est une variable aléatoire.

b. Vrai. L'estimateur du maximum de vraisemblance est une fonction des variables aléatoires de l'échantillon.

c. Faux. Au sens strict, la solution du programme de maximisation de la log-vraisemblance correspond à une réalisation de l'estimateur du maximum de vraisemblance (estimation). On en déduit ensuite un estimateur (variable aléatoire).

d. Vrai. Le gradient de l'échantillon, évalué au point de la réalisation de l'estimateur du maximum de vraisemblance, est égal à zéro.

e. Faux. La hessienne associée à l'échantillon est alors une matrice de dimension 3×3.

3 Score, hessienne et information de Fisher

a. Vrai. On peut définir le gradient comme une réalisation du score.

b. Vrai. Par définition, l'espérance du score est nulle quelle que soit valeur du paramètre.

c. Vrai. C'est l'une des deux définitions de l'information de Fisher associée à l'échantillon.

d. Vrai. Par construction, l'information de Fisher de l'échantillon est égale à l'information de Fisher moyenne multipliée par la taille de l'échantillon.

e. Faux. Ceci n'est vrai que dans le cas de lois marginales. Dans le cas de modèles économétriques (lois conditionnelles), l'information de Fisher moyenne correspond à l'espérance (par rapport aux variables explicatives X) de l'information de Fisher associée à une observation de l'échantillon.

4 Propriétés de l'estimateur du maximum de vraisemblance

a. Faux. L'estimateur du maximum de vraisemblance est sans biais, uniquement lorsque les hypothèses de régularité sont satisfaites.

b. Faux. Sous les hypothèses de régularité usuelles, l'estimateur du maximum de vraisemblance est convergent au sens faible.

c. Vrai. Sous les hypothèses de régularité, l'estimateur du maximum de vraisemblance est asymptotiquement normalement distribué.

d. Faux. Sa variance asymptotique est égale à la borne FDCR.

e. Faux. La variance asymptotique de l'estimateur du maximum de vraisemblance est égale à l'*inverse* de la matrice d'information de Fisher associée à l'échantillon.

5 Maximum de vraisemblance

1. Puisque le paramètre μ est connu, nous exprimons le logarithme de la densité de X en fonction uniquement du paramètre inconnu σ^2 :

$$\ln f_X\left(x;\sigma^2\right) = -\ln x - \frac{1}{2}\ln\left(\sigma^2\right) - \frac{1}{2}\ln(2\pi) - \frac{(\ln x-\mu)^2}{2\sigma^2}$$

Puisque les variables $X_1,\ldots,X_n$ sont indépendantes, la log-vraisemblance de l'échantillon $(x_1,\ldots,x_n)$ est :

$$\ell_n\left(\sigma^2;x\right) = \ln L_n\left(\sigma^2;x\right) = \sum_{i=1}^{n}\ln f_X\left(x_i;\sigma^2\right)$$

On obtient :

$$\ell_n\left(\sigma^2;x\right) = -\sum_{i=1}^{n}\ln x_i - \frac{n}{2}\ln\left(\sigma^2\right) - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln x_i-\mu)^2$$

2. L'estimateur du maximum de vraisemblance est défini par :

$$\widehat{\sigma}^2 = \underset{\sigma^2\in\mathbb{R}^+}{\arg\max}\ \ell_n\left(\sigma^2;x\right)$$

Pour simplifier la dérivation de la log-vraisemblance, on pose $\theta = \sigma^2$.

$$\ell_n\left(\sigma^2;x\right) = -\sum_{i=1}^{n}\ln x_i - \frac{n}{2}\ln(\theta) - \frac{n}{2}\ln(2\pi) - \frac{1}{2\theta}\sum_{i=1}^{n}(\ln x_i-\mu)^2$$

La condition du premier ordre (équation de log-vraisemblance) est :

$$\text{CN} : g_n\left(\widehat{\theta},x\right) = \left.\frac{\partial\ell_n\left(\sigma^2;x\right)}{\partial\theta}\right|_{\widehat{\theta}} = -\frac{n}{2\widehat{\theta}} + \frac{1}{2\widehat{\theta}^2}\sum_{i=1}^{n}(\ln x_i-\mu)^2 = 0$$

D'où l'on tire que :

$$\widehat{\theta} = \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\ln x_i-\mu)^2$$

On vérifie qu'il s'agit d'un maximum :

$$\text{CS} : H_n\left(\widehat{\theta},x\right) = \left.\frac{\partial^2 \ell_n\left(\sigma^2;x\right)}{\partial\theta^2}\right|_{\widehat{\theta}} = \frac{n}{2\widehat{\theta}^2} - \frac{1}{\widehat{\theta}^3}\sum_{i=1}^{n}(\ln x_i - \mu)^2$$

Sachant que $\sum_{i=1}^{n}(\ln(x_i) - \mu)^2 = n \times \widehat{\theta}$, on obtient :

$$\left.\frac{\partial^2 \ell_n\left(\sigma^2;x\right)}{\partial\theta^2}\right|_{\widehat{\theta}} = \frac{n}{2\widehat{\theta}^2} - \frac{n\times\widehat{\theta}}{\widehat{\theta}^3} = -\frac{n}{2\widehat{\theta}^2} = -\frac{n}{2\widehat{\sigma}^4} < 0$$

On a bien un maximum. Par conséquent, l'estimateur du maximum de vraisemblance $\widehat{\sigma}^2$ du paramètre σ^2 est défini par :

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\ln X_i - \mu)^2$$

3. Calculons l'espérance de l'estimateur $\widehat{\sigma}^2$:

$$\mathbb{E}\left(\widehat{\sigma}^2\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}(\ln X_i - \mu)^2\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left((\ln X_i - \mu)^2\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left((\ln X_i - \mathbb{E}(\ln X_i))^2\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{V}(\ln X_i)$$

Puisque $\mathbb{V}(\ln X_i) = \sigma^2$, nous obtenons :

$$\mathbb{E}\left(\widehat{\sigma}^2\right) = \frac{n\times\sigma^2}{n} = \sigma^2$$

L'estimateur $\widehat{\sigma}^2$ est sans biais.

4. Il y a plusieurs façons de montrer ce résultat. Une façon consiste à utiliser la loi faible des grands nombres (▸ chapitre 8). La séquence de variables $(\ln X_i - \mu)^2$ est i.i.d. avec une espérance égale à :

$$\mathbb{E}\left((\ln X_i - \mu)^2\right) = \mathbb{E}\left((\ln X_i - \mathbb{E}(\ln X_i))^2\right) = \mathbb{V}(\ln X_i) = \sigma^2$$

D'après la loi faible des grands nombres, la moyenne empirique de variables i.i.d. converge en probabilité vers l'espérance :

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\ln X_i - \mu)^2 \xrightarrow{p} \mathbb{E}\left((\ln X_i - \mu)^2\right) = \sigma^2$$

L'estimateur $\widehat{\sigma}^2$ est convergent au sens faible.

5. Le score associé à l'échantillon est défini par :

$$S_n\left(\sigma^2,X\right) = \frac{\partial \ell_n\left(\sigma^2;x\right)}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(\ln X_i - \mu)^2$$

Par conséquent :

$$\mathbb{E}\left(S_n\left(\sigma^2,X\right)\right) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}\mathbb{E}\left((\ln X_i - \mu)^2\right)$$

$$= -\frac{n}{2\sigma^2} + \frac{n\times\sigma^2}{2\sigma^4} = 0 \quad \forall\sigma^2 \in \mathbb{R}^+$$

6. La quantité d'information de Fisher associée à l'échantillon est égale à :

$$\mathrm{I}_n\left(\sigma^2\right) = \mathbb{E}\left(-H_n\left(\sigma^2;X\right)\right)$$

$$= \mathbb{E}\left(-\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{i=1}^{n}(\ln X_i - \mu)^2\right)$$

$$= -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{i=1}^{n}\mathbb{E}\left((\ln X_i - \mu)^2\right)$$

$$= -\frac{n}{2\sigma^4} + \frac{n\times\sigma^2}{\sigma^6} = \frac{n}{2\sigma^4}$$

La quantité moyenne d'information de Fisher est égale à :

$$\mathrm{I}\left(\sigma^2\right) = \frac{1}{n}\times\mathrm{I}_n\left(\sigma^2\right) = \frac{1}{2\sigma^4}$$

7. La fonction $\ln f_X\left(x;\sigma^2\right)$ satisfait les hypothèses de régularité. Par conséquent, l'estimateur du maximum de vraisemblance est asymptotiquement normalement distribué :

$$\sqrt{n}\left(\widehat{\sigma}^2 - \sigma^2\right) \xrightarrow{d} \mathcal{N}\left(0,\mathrm{I}^{-1}\left(\sigma^2\right)\right)$$

où $\mathrm{I}\left(\sigma^2\right)$ correspond à la quantité moyenne d'information de Fisher. Ainsi, nous avons :

$$\sqrt{n}\left(\widehat{\sigma}^2 - \sigma^2\right) \xrightarrow{d} \mathcal{N}\left(0,2\sigma^4\right)$$

Pour une taille d'échantillon suffisamment grande mais finie, ce résultat implique que :

$$\widehat{\sigma}^2 \overset{asy}{\approx} \mathcal{N}\left(\sigma^2,\frac{2\sigma^4}{n}\right)$$

8. La variance asymptotique de l'estimateur $\widehat{\sigma}^2$ est égale à :

$$\mathbb{V}_{asy}\left(\widehat{\sigma}^2\right) = \frac{2\sigma^4}{n}$$

La borne FDCR est définie par l'inverse de la quantité d'information de Fisher associée à l'échantillon :

$$\mathrm{I}_n^{-1}\left(\sigma^2\right) = \frac{2\sigma^4}{n}$$

L'estimateur $\widehat{\sigma}^2$ est asymptotiquement efficace puisque :

$$\mathbb{V}_{asy}\left(\widehat{\sigma}^2\right) = \mathrm{I}_n^{-1}\left(\sigma^2\right) = \frac{2\sigma^4}{n}$$

9. Un estimateur convergent de cette variance asymptotique est donné par :

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\sigma}^2\right) = \frac{2\widehat{\sigma}^4}{n}$$

Un autre estimateur basé sur la hessienne (dérivée seconde de la log-vraisemblance) est donné par :

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\sigma}^2\right) = \left(-\sum_{i=1}^{n}H_i\left(\widehat{\theta},x_i\right)\right) = -\frac{n}{2\widehat{\theta}^2} + \frac{1}{\widehat{\theta}^3}\sum_{i=1}^{n}(\ln x_i - \mu)^2$$

Asymptotiquement, les deux estimateurs donnent la même valeur, mais à distance finie ils peuvent fournir des estimations différentes.

6 Maximum de vraisemblance

1. On sait que les variables $(X_1,\ldots,X_N)$ sont i.i.d. de même loi que X. Dès lors, la log-vraisemblance associée au n-échantillon $(X_1,..,X_N)$ s'écrit :

$$\begin{aligned}\ell_n(\beta;x) &= \ln L_n(\beta;x)\\ &= \sum_{i=1}^{n}\ln f_X(x_i;\alpha,\beta)\\ &= (\alpha-1)\sum_{i=1}^{n}\ln(x_i)-\frac{1}{\beta}\sum_{i=1}^{n}x_i\\ &\quad -n\ln(\Gamma(\alpha))-n\alpha\ln(\beta)\end{aligned}$$

2. Le gradient est défini par :

$$g_n(\beta;x)=\frac{\partial \ell_n(\beta;x)}{\partial\beta}=\frac{1}{\beta^2}\sum_{i=1}^{n}x_i-\frac{n\alpha}{\beta}$$

La hessienne est définie par :

$$H_n(\beta;x)=\frac{\partial^2\ell_n(\beta;x)}{\partial\beta^2}=-\frac{2}{\beta^3}\sum_{i=1}^{n}x_i+\frac{n\alpha}{\beta^2}$$

3. Le score associé à l'échantillon est égal à :

$$S_n(\beta;X)=\frac{\partial \ell_n(\beta;X)}{\partial\beta}=\frac{1}{\beta^2}\sum_{i=1}^{n}X_i-\frac{n\alpha}{\beta}$$

Son espérance est égale à :

$$\begin{aligned}\mathbb{E}(S_n(\beta;X)) &= \mathbb{E}\left(\frac{1}{\beta^2}\sum_{i=1}^{n}X_i-\frac{n\alpha}{\beta}\right)\\ &= \frac{1}{\beta^2}\sum_{i=1}^{n}\mathbb{E}(X_i)-\frac{n\alpha}{\beta}\\ &= \frac{n\alpha\beta}{\beta^2}-\frac{n\alpha}{\beta}=0\end{aligned}$$

4. Soit $\widehat{\beta}$ l'estimateur du maximum de vraisemblance du paramètre β. Ce dernier vérifie :

$$\widehat{\beta}=\arg\max_{\beta\in\mathbb{R}^+}\ell_n(\beta;x)$$

La condition nécessaire du programme d'optimisation de la log-vraisemblance (équation de log-vraisemblance) s'écrit :

$$g_n\left(\widehat{\beta};x\right)=\frac{1}{\widehat{\beta}^2}\sum_{i=1}^{n}x_i-\frac{n\alpha}{\widehat{\beta}}=0$$

D'où l'on tire que :

$$\widehat{\beta}(x)=\frac{1}{\alpha n}\sum_{i=1}^{n}x_i$$

La condition suffisante se déduit de $H_n\left(\widehat{\beta};x\right)$:

$$\begin{aligned}H_n\left(\widehat{\beta};x\right) &= -\frac{2}{\widehat{\beta}^3}\sum_{i=1}^{n}x_i+\frac{n\alpha}{\widehat{\beta}^2}\\ &= -\frac{2\alpha n\widehat{\beta}}{\widehat{\beta}^3}+\frac{n\alpha}{\widehat{\beta}^2}=-\frac{\alpha n}{\widehat{\beta}^2}<0\end{aligned}$$

puisque le paramètre α est positif et $\widehat{\beta}$ est positif (moyenne de réalisations x_i positives). L'estimateur du maximum de vraisemblance (variable aléatoire) est défini par :

$$\widehat{\beta}=\frac{1}{\alpha n}\sum_{i=1}^{n}X_i$$

5. Calculons $\mathbb{E}\left(\widehat{\beta}\right)$:

$$\mathbb{E}\left(\widehat{\beta}\right)=\mathbb{E}\left(\frac{1}{\alpha n}\sum_{i=1}^{n}X_i\right)=\frac{1}{\alpha n}\sum_{i=1}^{n}\mathbb{E}(X_i)=\frac{n\alpha\beta}{\alpha n}=\beta$$

L'estimateur $\widehat{\beta}$ est sans biais.

6. Puisque les X_i sont i.i.d. de même loi que X, il vient :

$$\mathbb{V}\left(\widehat{\beta}\right)=\mathbb{V}\left(\frac{1}{\alpha n}\sum_{i=1}^{n}X_i\right)=\frac{1}{\alpha^2n^2}\sum_{i=1}^{n}\mathbb{V}(X_i)=\frac{n\alpha\beta^2}{\alpha^2n^2}=\frac{\beta^2}{\alpha n}$$

Par conséquent :

$$\lim_{n\to\infty}\mathbb{V}\left(\widehat{\beta}\right)=\lim_{n\to\infty}\frac{\beta^2}{\alpha n}=0$$

Sachant que par ailleurs $\mathbb{E}\left(\widehat{\beta}\right)=\beta$, on en déduit que $\widehat{\beta}$ est un estimateur convergent (au sens faible) de β :

$$\widehat{\beta}\xrightarrow{p}\beta$$

7. La quantité d'information de Fisher associée à l'échantillon est définie par :

$$\mathrm{I}_n(\beta)=\mathbb{V}(S_n(\beta;X))=\mathbb{E}(-H_n(\beta;X))$$

D'après les résultats de la question 2, il vient :

$$\begin{aligned}\mathrm{I}_n(\beta) &= \mathbb{E}\left(\frac{2}{\beta^3}\sum_{i=1}^{n}X_i-\frac{n\alpha}{\beta^2}\right)=\frac{2}{\beta^3}\sum_{i=1}^{n}\mathbb{E}(X_i)-\frac{n\alpha}{\beta^2}\\ &= \frac{2n\alpha\beta}{\beta^3}-\frac{n\alpha}{\beta^2}=\frac{n\alpha}{\beta^2}\end{aligned}$$

Dès lors :

$$\mathbb{V}\left(\widehat{\beta}\right)=\mathrm{I}_n^{-1}(\beta)=\frac{\beta^2}{\alpha n}$$

L'estimateur du maximum de vraisemblance $\widehat{\beta}$ est efficace au sens de la borne FDCR.

8. Les variables du n-échantillon $(X_1,\ldots,X_n)$ sont i.i.d. avec $\mathbb{E}(X_i)=\alpha\beta$ et $\mathbb{V}(X_i)=\alpha\beta^2$. D'après le théorème central limite :

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}X_i-\mathbb{E}(X_i)\right)\xrightarrow{d}\mathcal{N}(0,\mathbb{V}(X))$$

ou encore :

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}X_i-\alpha\beta\right)\xrightarrow{d}\mathcal{N}\left(0,\alpha\beta^2\right)$$

On en déduit la loi asymptotique de l'estimateur $\widehat{\beta}=(\alpha n)^{-1}\sum_{i=1}^{n}X_i$ par la méthode delta en posant $\widehat{\beta}=g\left(\overline{X}_n\right)$ avec $g(x)=x/\alpha$ et $\partial g(x)/\partial x=1/\alpha$.

$$\sqrt{n}\left(g\left(\overline{X}_n\right)-g(\alpha\beta)\right)\xrightarrow{d}\mathcal{N}\left(0,\alpha\beta^2\left(\left.\frac{\partial g(x)}{\partial x}\right|_{\alpha\beta}\right)^2\right)$$

On obtient finalement :

$$\sqrt{n}\left(\widehat{\beta}-\beta\right)\xrightarrow{d}\mathcal{N}\left(0,\frac{\beta^2}{\alpha}\right)$$

9. Le problème est régulier, dès lors :

$$\sqrt{n}\left(\widehat{\beta}-\beta\right) \xrightarrow{d} \mathcal{N}\left(0,\mathrm{I}(\beta)^{-1}\right)$$

où $\mathrm{I}(\beta)$ désigne la quantité d'information moyenne de Fisher, définie par :

$$\mathrm{I}(\beta) = \frac{1}{n}\mathrm{I}_n(\beta) = \frac{\alpha}{\beta^2}$$

On en conclut que :

$$\sqrt{n}\left(\widehat{\beta}-\beta\right) \xrightarrow{d} \mathcal{N}\left(0,\frac{\beta^2}{\alpha}\right)$$

On retrouve la même loi asymptotique qu'à la question précédente.

10. On sait que pour $\alpha = 2$:

$$\widehat{\beta} = \frac{1}{\alpha n}\sum_{i=1}^{n} X_i = \frac{\overline{X}_n}{2}$$

Sachant que dans cet échantillon, on obtient $\overline{x}_N = 4$, la réalisation de l'estimateur du maximum de vraisemblance est égale à :

$$\widehat{\beta}(x) = 2$$

Chapitre 11

1 Règle de décision d'un test statistique

a. Faux. Un test statistique est une règle de décision qui porte sur le rejet ou le non rejet de l'hypothèse nulle.
b. Faux. Le rejet de l'hypothèse nulle n'implique pas nécessairement l'acceptation de l'hypothèse alternative.
c. Faux. Même si l'on ne rejette pas l'hypothèse nulle, cela ne signifie pas nécessairement que l'on accepte l'hypothèse nulle ou que cette hypothèse est vraie.
d. Vrai. C'est précisément ce que l'on peut conclure d'un test.
e. Faux. Le rejet ou le non rejet de l'hypothèse nulle dépend du niveau du test.

2 Région critique d'un test

a. Vrai. C'est la définition de la région critique.
b. Vrai. Comme toute statistique, une statistique de test est une fonction des variables aléatoires de l'échantillon, c'est donc une variable aléatoire.
c. Faux. Ce n'est pas toujours le cas. On peut définir un test à partir d'une transformée d'un estimateur du paramètre dont on connaît la loi sous l'hypothèse nulle.
d. Vrai.
e. Vrai. C'est précisément la règle de décision d'un test.

3 Niveau et puissance d'un test

a. Faux. Le niveau d'un test correspond à la probabilité de rejeter à tort l'hypothèse nulle.
b. Faux. Cela dépend de la forme de la région critique.
c. Vrai. C'est ainsi que l'on résout l'arbitrage entre le risque de première espèce et le risque de deuxième espèce en fonction de la valeur critique du test.
d. Vrai. C'est la définition de la puissance qui correspond à un moins la probabilité de risque de deuxième espèce.
e. Faux. C'est la puissance d'un test convergent qui tend vers l'unité lorsque la taille d'échantillon tend vers l'infini.

4 Test paramétrique et lemme de Neyman-Pearson

a. Vrai. C'est la définition de l'objet du lemme de Neyman-Pearson.
b. Faux. La région critique sera la forme $W=\{x : \widehat{\theta}(x) < c\}$. C'est lorsque la réalisation $\widehat{\theta}(x)$ est relativement petite, par rapport à la valeur critique, que l'on rejette l'hypothèse nulle.
c. Faux. La région critique est de la forme $W=\{x : |\widehat{\theta}(x) - \theta_0| < c\}$. Lorsque la distance $|\widehat{\theta}(x) - \theta_0|$ (en valeur absolue) excède la valeur critique, on rejette l'hypothèse nulle.
d. Faux. Il n'existe pas de test UPP dans le cas d'un test bilatéral.
e. Faux. C'est la région de non rejet d'un test bilatéral de niveau α qui correspond à l'intersection des régions de non rejet des tests unilatéraux associés de niveau $\alpha/2$.

5 Tests paramétriques

1. La log-vraisemblance associée à l'échantillon $(x_1,\ldots,x_n)$ s'écrit :

$$\ell_n\left(\sigma^2;x\right) = \ln L_n\left(\sigma^2;x\right) = \sum_{i=1}^{n}\ln(x_i) - n\ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2$$

D'après le lemme de Neyman-Pearson, la région critique du test UPP de niveau α est déterminée par :

$$W = \left\{ x \middle| \frac{L_n\left(\sigma_0^2;x\right)}{L_n\left(\sigma_1^2;x\right)} < k \right\}$$

où k est une constante déterminée par le niveau de risque de première espèce α. Cette inégalité peut se réécrire sous la forme :

$$\ell_n\left(\sigma_0^2;x\right) - \ell_n\left(\sigma_1^2;x\right) < \ln(k)$$

$$\Longleftrightarrow n\left(\log\left(\sigma_1^2\right) - \log\left(\sigma_0^2\right)\right) + \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\frac{1}{2}\sum_{i=1}^{n}x_i^2 < \ln(k)$$

$$\Longleftrightarrow \left(\frac{\sigma_0^2 - \sigma_1^2}{\sigma_0^2\sigma_1^2}\right)\frac{1}{2}\sum_{i=1}^{n}x_i^2 < k_1$$

avec $k_1 = \ln(k) - n\left(\log\left(\sigma_1^2\right) - \log\left(\sigma_0^2\right)\right)$. Puisque dans notre cas $\sigma_1^2 > \sigma_0^2$, on a :

$$\frac{1}{2n}\sum_{i=1}^{n}x_i^2 > c$$

où $c = k_1\sigma_0^2\sigma_1^2/\left(\left(\sigma_0^2 - \sigma_1^2\right)n\right)$. La région critique du test UPP de niveau α est de la forme :

$$W = \{x | T_n(x) > c\}$$

où c est une constante déterminée par le risque de première espèce.

2. La distribution asymptotique de la statistique de test T_n sous l'hypothèse nulle $H_0 : \sigma^2 = \sigma_0^2$ est :

$$T_n \underset{H_0}{\overset{asy}{\approx}} \mathcal{N}\left(\sigma_0^2, \frac{\sigma_0^4}{n}\right) \Longleftrightarrow \frac{T_n - \sigma_0^2}{\sigma_0^2/\sqrt{n}} \underset{H_0}{\overset{asy}{\approx}} \mathcal{N}(0,1)$$

Par définition du risque de première espèce :

$$\begin{aligned} \alpha &= \Pr(W|H_0) = \Pr(T_n > c|H_0) \\ &= 1 - \Pr\left(\frac{T_n - \sigma_0^2}{\sigma_0^2/\sqrt{n}} \leq \frac{c - \sigma_0^2}{\sigma_0^2/\sqrt{n}}\middle| H_0\right) \\ &= 1 - \Phi\left(\frac{c - \sigma_0^2}{\sigma_0^2/\sqrt{n}}\right) \end{aligned}$$

où $\Phi(.)$ désigne la fonction de répartition de la loi normale centrée réduite. D'où l'on tire que :

$$1 - \alpha = \Phi\left(\frac{c - \sigma_0^2}{\sigma_0^2/\sqrt{n}}\right)$$

La valeur critique associée à un niveau $\alpha = 5\ \%$ est égale à :

$$c = \sigma_0^2 + \Phi^{-1}(1 - \alpha)\frac{\sigma_0^2}{\sqrt{n}} \qquad (*)$$

Il vient :

$$c = 2 + \Phi^{-1}(0{,}95) \times \frac{2}{\sqrt{500}} = 2 + 1{,}64 \times \frac{2}{\sqrt{500}} = 2{,}14$$

La région critique du test UPP de niveau $\alpha = 5\ \%$ est définie par :

$$W = \{x|\, T_n(x) > 2{,}14\}$$

3. Sous l'hypothèse alternative $H_1 : \sigma^2 = \sigma_1^2 = 2{,}1$, la distribution asymptotique de la statistique de test T_n devient :

$$T_n \underset{H_1}{\overset{asy}{\approx}} \mathcal{N}\left(\sigma_1^2, \frac{\sigma_1^4}{n}\right) \Longleftrightarrow \frac{T_n - \sigma_1^2}{\sigma_1^2/\sqrt{n}} \underset{H_1}{\overset{asy}{\approx}} \mathcal{N}(0,1)$$

Par définition de la puissance, $\Pr(W|H_1)$, on a :

$$\begin{aligned} \text{Puissance} &= \Pr(T_n > c|H_1) \\ &= 1 - \Pr\left(\frac{T_n - \sigma_1^2}{\sigma_1^2/\sqrt{n}} \leq \frac{c - \sigma_1^2}{\sigma_1^2/\sqrt{n}}\middle| H_1\right) \\ &= 1 - \Phi\left(\frac{c - \sigma_1^2}{\sigma_1^2/\sqrt{n}}\right) \qquad (**) \end{aligned}$$

On a donc une puissance égale à :

$$\begin{aligned} \text{Puissance} &= 1 - \Phi\left(\frac{2{,}14 - 2{,}1}{2{,}1/\sqrt{500}}\right) = 1 - \Phi(0{,}42) \\ &= 1 - 0{,}66 = 0{,}34 \end{aligned}$$

Avec cette règle de décision, il y a 34 % de chances de rejeter l'hypothèse $H_0 : \sigma^2 = \sigma_0^2 = 2$ lorsque l'hypothèse $H_1 : \sigma^2 = \sigma_1^2 = 2{,}1$ est vraie dans la population.

4. En remplaçant c par son expression (équation (*)) dans la puissance (équation (**)), il vient :

$$\text{Puissance} = 1 - \Phi\left(\frac{\sigma_0^2 - \sigma_1^2}{\sigma_1^2/\sqrt{n}} + \frac{\sigma_0^2}{\sigma_1^2}\Phi^{-1}(1 - \alpha)\right)$$

On vérifie que si σ_1^2 tend vers σ_0^2, alors la puissance tend vers le risque de première espèce α (le test est non biaisé).

De plus, puisque $\sigma_0^2 < \sigma_1^2$, si n tend vers l'infini, la puissance tend vers l'unité. En effet :

$$\lim_{n\to\infty} \frac{\sigma_0^2 - \sigma_1^2}{\sigma_1^2/\sqrt{n}} = -\infty$$

Par conséquent, $\forall \sigma_1^2 > \sigma_0^2$, on a :

$$\begin{aligned} 1 - \lim_{n\to\infty} \Phi\left(\frac{\sigma_0^2 - \sigma_1^2}{\sigma_1^2/\sqrt{n}} + \frac{\sigma_0^2}{\sigma_1^2}\Phi^{-1}(1 - \alpha)\right) &= 1 - \Phi(-\infty) \\ &= 1 - 0 = 1 \end{aligned}$$

Le test est convergent.

5. La réalisation de la statistique de test $T_n(x)$ est égale à :

$$T_n(x) = \frac{1}{2n}\sum_{i=1}^{n} x_i^2 = \frac{2\,070}{2 \times 500} = 2{,}07$$

La région critique du test pour un niveau $\alpha = 5\ \%$ est $W = \{x|\, T_n(x) > 2{,}14\}$. La réalisation de la statistique de test n'appartient pas à la région critique. Par conséquent, on ne peut pas rejeter l'hypothèse nulle $H_0 : \sigma^2 = 2$ pour un seuil de risque de 5 %.

6. La région critique du test $H_0 : \sigma^2 = \sigma_0^2$ contre $H_1 : \sigma^2 = \sigma_1^2$ ne dépend pas de la valeur de σ_1^2. La région critique du test UPP unilatéral $H_0 : \sigma^2 = \sigma_0^2$ contre $H_1 : \sigma^2 > \sigma_0^2$ est identique :

$$\begin{aligned} W &= \left\{x|\, T_n(x) > \sigma_0^2 + \Phi^{-1}(1 - \alpha)\frac{\sigma_0^2}{\sqrt{n}}\right\} \\ &= \{x|\, T_n(x) > 2{,}14\} \end{aligned}$$

7. La région de non rejet du test bilatéral de niveau α est définie par l'intersection des régions de non rejet des tests unilatéraux UPP de niveau $\alpha/2$:

Test A : $H_0 : \sigma^2 = \sigma_0^2$ contre $H_1 : \sigma^2 > \sigma_0^2$

Test B : $H_0 : \sigma^2 = \sigma_0^2$ contre $H_1 : \sigma^2 < \sigma_0^2$

Soient $\overline{W}_A$ et $\overline{W}_B$ les régions de non rejet associées :

$$\begin{aligned} \overline{W}_A &= \left\{x|\, T_n(x) < \sigma_0^2 + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\frac{\sigma_0^2}{\sqrt{n}}\right\} \\ &= \{x|\, T_n(x) < 2{,}17\} \end{aligned}$$

$$\begin{aligned} \overline{W}_B &= \left\{x|\, T_n(x) > \sigma_0^2 + \Phi^{-1}\left(\frac{\alpha}{2}\right)\frac{\sigma_0^2}{\sqrt{n}}\right\} \\ &= \{x|\, T_n(x) > 1{,}83\} \end{aligned}$$

La région de non rejet $\overline{W} = \overline{W}_A \cap \overline{W}_B$ et la région critique W du test bilatéral de niveau $\alpha = 5\ \%$ sont donc respectivement définies par :

$$\overline{W} = \{x|\, 1{,}83 < T_n(x) < 2{,}17\}$$
$$W = \{x|\, T_n(x) \notin [1{,}83; 2{,}17]\}$$

De façon générale, la région de non rejet peut aussi s'écrire sous la forme :

$$\overline{W} = \left\{x\middle| \left|\frac{T_n(x) - \sigma_0^2}{\sigma_0^2/\sqrt{n}}\right| < \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right\}$$

La région critique du test bilatéral de niveau α est alors définie par :

$$W = \left\{ x \left| \left| \frac{T_n(x) - \sigma_0^2}{\sigma_0^2/\sqrt{n}} \right| \geq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \right. \right\}$$

6 Tests paramétriques

1. Les variables $X_1,\ldots,X_n$ sont i.i.d., par conséquent la log-vraisemblance associée à l'échantillon $(x_1,\ldots,x_n)$ est égale à :

$$\ell_n(\theta; x) = \sum_{i=1}^{n} \ln f_X(x_i;\theta)$$

avec :

$$\ln f_X(x_i;\theta) = -\ln(\theta) - \frac{1}{\theta}\ln(c) + \left(\frac{1-\theta}{\theta}\right)\ln(x_i)$$

Ainsi, on obtient :

$$\ell_n(\theta; x) = -n\ln(\theta) - \frac{n}{\theta}\ln(c) + \left(\frac{1-\theta}{\theta}\right)\sum_{i=1}^{n}\ln(x_i)$$

2. L'estimateur du maximum de vraisemblance $\widehat{\theta}$ du paramètre θ vérifie :

$$\widehat{\theta} = \underset{\theta\in\mathbb{R}^+}{\arg\max}\ \ell_n(\theta; x)$$

La condition nécessaire (équation de log-vraisemblance) est :

$$g_n\left(\widehat{\theta}; x\right) = \left.\frac{\partial \ell_n(\theta; x)}{\partial\theta}\right|_{\widehat{\theta}}$$
$$= -\frac{n}{\widehat{\theta}} + \frac{n}{\widehat{\theta}^2}\ln(c) - \frac{1}{\widehat{\theta}^2}\sum_{i=1}^{n}\ln(x_i) = 0$$

En résolvant cette équation, il vient :

$$\widehat{\theta} = \ln(c) - \frac{1}{n}\sum_{i=1}^{n}\ln(x_i)$$

La condition suffisante de ce programme est :

$$H_n\left(\widehat{\theta}; x\right) = \left.\frac{\partial^2 \ell_n(\theta; x)}{\partial\theta^2}\right|_{\widehat{\theta}} = \frac{n}{\widehat{\theta}^2} - \frac{2n}{\widehat{\theta}^3}\left(\ln(c) - \frac{1}{n}\sum_{i=1}^{n}\ln(x_i)\right)$$

En utilisant les résultats de la condition nécessaire, il vient :

$$H_n\left(\widehat{\theta}; x\right) = \frac{n}{\widehat{\theta}^2} - \frac{2n}{\widehat{\theta}^3}\left(\ln(c) - \frac{1}{n}\sum_{i=1}^{n}\ln(x_i)\right)$$
$$= \frac{n}{\widehat{\theta}^2} - \frac{2n\widehat{\theta}}{\widehat{\theta}^3}$$
$$= -\frac{n}{\widehat{\theta}^2} < 0$$

Nous avons donc un maximum. L'estimateur (variable aléatoire) du maximum de vraisemblance $\widehat{\theta}$ du paramètre θ est égal à :

$$\widehat{\theta} = \ln(c) - \frac{1}{n}\sum_{i=1}^{n}\ln(X_i)$$

3. La séquence de variables aléatoires i.i.d. $\ln(X_1),\ldots,\ln(X_n)$ vérifie $\mathbb{E}(\ln(X_i)) = \ln(c) - \theta$. D'après la loi faible des grands nombres :

$$\frac{1}{n}\sum_{i=1}^{n}\ln(X_i) \xrightarrow{p} \mathbb{E}(\ln(X_i)) = \ln(c) - \theta$$

En utilisant le *continuous mappin theorem* pour une fonction $g(z) = \ln(c) - z$, on obtient :

$$g\left(\frac{1}{n}\sum_{i=1}^{n}\ln(X_i)\right) \xrightarrow{p} g(\ln(c) - \theta)$$

ou de façon équivalente :

$$\ln(c) - \frac{1}{n}\sum_{i=1}^{n}\ln(X_i) \xrightarrow{p} \ln(c) - \ln(c) + \theta$$

Par conséquent :

$$\widehat{\theta} \xrightarrow{p} \theta$$

L'estimateur $\widehat{\theta}$ est convergent (au sens faible).

4. Puisque le problème est régulier, nous avons :

$$\sqrt{n}\left(\widehat{\theta} - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \mathrm{I}^{-1}(\theta)\right)$$

où θ désigne la vraie valeur du paramètre et $\mathrm{I}(\theta_0)$ correspond à la quantité d'information de Fisher moyenne.

$$\mathrm{I}(\theta) = \frac{1}{n} \times \mathrm{I}_n(\theta)$$

avec :

$$\mathrm{I}_n(\theta) = \mathbb{E}(-H_n(\theta; X))$$
$$= \mathbb{E}\left(-\frac{n}{\theta^2} + \frac{2n}{\theta^3}\left(\ln(c) - \frac{1}{n}\sum_{i=1}^{n}\ln(X_i)\right)\right)$$
$$= -\frac{n}{\theta^2} + \frac{2n}{\theta^3}\left(\ln(c) - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(\ln(X_i))\right)$$
$$= -\frac{n}{\theta^2} + \frac{2n}{\theta^3}(\ln(c) - \ln(c) + \theta)$$
$$= -\frac{n}{\theta^2} + \frac{2n\theta}{\theta^3}$$
$$= \frac{n}{\theta^2}$$

Par conséquent :

$$\sqrt{n}\left(\widehat{\theta} - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \theta^2\right)$$

ou de façon équivalente :

$$\widehat{\theta} \overset{asy}{\approx} \mathcal{N}\left(\theta, \frac{\theta^2}{n}\right)$$

5. D'après le lemme de Neyman Pearson, la région critique du test UPP de niveau α est donné par :

$$\frac{L_n(\theta_0; x)}{L_n(\theta_1; x)} < K$$

où K est une constante déterminée par le niveau α ou de façon équivalente :

$$\ell_n(\theta_0; x) - \ell_n(\theta_1; x) < \ln(K)$$

On obtient :

$$-n\ln(\theta_0) \quad - \frac{n}{\theta_0}\ln(c) + \left(\frac{1}{\theta_0} - 1\right)\sum_{i=1}^{n}\ln(x_i) + n\ln(\theta_1)$$
$$+ \frac{n}{\theta_1}\ln(c) - \left(\frac{1}{\theta_1} - 1\right)\sum_{i=1}^{n}\ln(x_i) < \ln(K)$$
$$\Longleftrightarrow \left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)\sum_{i=1}^{n}\ln(x_i) < K_2$$

où $K_2 = \ln(K) + n(\ln(\theta_0) - \ln(\theta_1)) + n\ln(c)\left(\theta_0^{-1} - \theta_1^{-1}\right)$ est un terme constant.

$$\left(\frac{\theta_1 - \theta_0}{\theta_0\theta_1}\right)\sum_{i=1}^{n}\ln(x_i) < K_2$$

Puisque $\theta_1 < \theta_0$, on a :

$$\sum_{i=1}^{n}\ln(x_i) > K_3$$

avec $K_3 = K_2\theta_0\theta_1/(\theta_1 - \theta_0)$. Cette inégalité peut se réécrire :

$$\ln(c) - \frac{1}{n}\sum_{i=1}^{n}\ln(x_i) < K_4$$

avec $K_4 = \ln(c) - K_3/n$. La région critique du test UPP de niveau α est de la forme :

$$\mathrm{W} = \left\{x : \widehat{\theta}(x) < A\right\}$$

où A est une constante déterminée par le niveau α.

6. D'après la définition du risque de première espèce :

$$\alpha = \Pr(\mathrm{W}|\,\mathrm{H}_0)$$

Donc :

$$\alpha = \Pr\left(\widehat{\theta} < A \middle| \widehat{\theta} \overset{asy}{\approx} \mathcal{N}\left(\theta_0, \frac{\theta_0^2}{n}\right)\right)$$

ou de façon équivalente :

$$\alpha = \Pr\left(\frac{\widehat{\theta} - \theta_0}{\theta_0/\sqrt{n}} < \frac{A - \theta_0}{\theta_0/\sqrt{n}} \middle| \frac{\widehat{\theta} - \theta_0}{\theta_0/\sqrt{n}} \overset{asy}{\approx} \mathcal{N}(0,1)\right)$$
$$= \Phi\left(\frac{A - \theta_0}{\theta_0/\sqrt{n}}\right)$$

où $\Phi(.)$ désigne la fonction de répartition de la loi normale centrée réduite. A partir de cette expression, on peut déduire la valeur critique du test UPP de niveau α :

$$A = \theta_0 + \frac{\theta_0}{\sqrt{n}}\Phi^{-1}(\alpha)$$

7. On considère le test d'hypothèse simple contre hypothèse simple :

$$\mathrm{H}_0 : \theta = \theta_0 \text{ contre } \mathrm{H}_1 : \theta = \theta_1$$

avec $\theta_1 < \theta_0$. La région critique du test UPP de niveau α est :

$$\mathrm{W} = \left\{x : \widehat{\theta}(x) < \theta_0 + \frac{\theta_0}{\sqrt{n}}\Phi^{-1}(\alpha)\right\}$$

Cette région critique ne dépend pas de la valeur de θ_1. Elle correspond donc à celle du test unilatéral :

$$\mathrm{H}_0 : \theta = \theta_0 \text{ contre } \mathrm{H}_1 : \theta < \theta_0$$

8. On considère les tests unilatéraux :

Test A : $\mathrm{H}_0 : \theta = \theta_0$ contre $\mathrm{H}_1 : \theta < \theta_0$

Test B : $\mathrm{H}_0 : \theta = \theta_0$ contre $\mathrm{H}_1 : \theta > \theta_0$

Les régions de non rejet des tests unilatéraux UPP de niveau $\alpha/2$ sont égales à :

$$\overline{\mathrm{W}}_{\mathrm{A}} = \left\{x : \widehat{\theta}(x) > \theta_0 + \frac{\theta_0}{\sqrt{n}}\Phi^{-1}\left(\frac{\alpha}{2}\right)\right\}$$
$$\overline{\mathrm{W}}_{\mathrm{B}} = \left\{x : \widehat{\theta}(x) < \theta_0 + \frac{\theta_0}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right\}$$

Donc la région de non rejet du test bilatéral de niveau α est définie par :

$$\overline{\mathrm{W}} = \left\{x : \theta_0 + \frac{\theta_0}{\sqrt{n}}\Phi^{-1}\left(\frac{\alpha}{2}\right) < \widehat{\theta}(x) < \theta_0 + \frac{\theta_0}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right\}$$

Puisque $\Phi^{-1}(\alpha/2) = -\Phi^{-1}(1 - \alpha/2)$, cette région peut se réécrire sous la forme :

$$\overline{\mathrm{W}} = \left\{x : \left|\widehat{\theta}(x) - \theta_0\right| < \frac{\theta_0}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right\}$$

La région critique du test bilatéral de niveau α est donnée par :

$$\mathrm{W} = \left\{x : \left|\widehat{\theta}(x) - \theta_0\right| > \frac{\theta_0}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right\}$$

9. Par définition de la fonction puissance :

$$\mathrm{P}(\theta) = \Pr(\mathrm{W}|\,\mathrm{H}_1) \quad \forall\theta \neq \theta_0$$

Sous l'hypothèse alternative :

$$\widehat{\theta} \underset{\mathrm{H}_1}{\overset{asy}{\approx}} \mathcal{N}\left(\theta, \frac{\theta^2}{n}\right)$$

avec $\theta \neq \theta_0$. La fonction puissance est égale à :

$$\mathrm{P}(\theta) = 1 - \Pr\left(\overline{\mathrm{W}} \middle| \mathrm{H}_1\right)$$
$$= 1 - \Pr\left(\theta_0 + \frac{\theta_0}{\sqrt{n}}\Phi^{-1}\left(\frac{\alpha}{2}\right) < \widehat{\theta} < \theta_0 + \frac{\theta_0}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right)$$
$$= 1 - \Pr\left(\widehat{\theta} < \theta_0 + \frac{\theta_0}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right)$$
$$+ \Pr\left(\widehat{\theta} < \theta_0 + \frac{\theta_0}{\sqrt{n}}\Phi^{-1}\left(\frac{\alpha}{2}\right)\right)$$
$$= 1 - \Phi\left(\frac{\theta_0 + \frac{\theta_0}{\sqrt{N}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) - \theta}{\theta/\sqrt{n}}\right)$$
$$+ \Phi\left(\frac{\theta_0 + \frac{\theta_0}{\sqrt{N}}\Phi^{-1}\left(\frac{\alpha}{2}\right) - \theta}{\theta/\sqrt{n}}\right)$$

On obtient :

$$P(\theta) = 1 - \Phi\left(\frac{\theta_0 - \theta}{\theta/\sqrt{n}} + \frac{\theta_0}{\theta}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) + \Phi\left(\frac{\theta_0 - \theta}{\theta/\sqrt{n}} + \frac{\theta_0}{\theta}\Phi^{-1}\left(\frac{\alpha}{2}\right)\right)$$

Lorsque n tend vers l'infini, deux cas doivent être considérés. Si $\theta > \theta_0$, alors :

$$\lim_{n\to\infty} P(\theta) = 1 - \Phi(-\infty) + \Phi(-\infty) = 1$$

Si $\theta < \theta_0$, on a :

$$\lim_{n\to\infty} P(\theta) = 1 - \Phi(+\infty) + \Phi(+\infty) = 1 - 1 + 1 = 1$$

Quelle que soit la valeur de θ, la fonction puissance tend vers 1 :

$$\lim_{n\to\infty} P(\theta) = 1$$

Le test est convergent.

Tables Statistiques Usuelles

Table 1: Table de la loi binomiale

$$P(X = k) = C_n^k p^k (1-p)^{n-k}$$

(k le nombre d'occurrences parmi n)

n = 10

k \ p	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
0	0,5987	0,3487	0,1969	0,1074	0,0563	0,0282	0,0135	0,0060	0,0025	0,0010
1	0,9139	0,7361	0,5443	0,3758	0,2440	0,1493	0,0860	0,0464	0,0233	0,0107
2	0,9885	0,9298	0,8202	0,6778	0,5256	0,3828	0,2616	0,1673	0,0996	0,0547
3	0,9990	0,9872	0,9500	0,8791	0,7759	0,6496	0,5138	0,3823	0,2660	0,1719
4	0,9999	0,9984	0,9901	0,9672	0,9219	0,8497	0,7515	0,6331	0,5044	0,3770
5	1,0000	0,9999	0,9986	0,9936	0,9803	0,9527	0,9051	0,8338	0,7384	0,6230
6	1,0000	1,0000	0,9999	0,9991	0,9965	0,9894	0,9740	0,9452	0,8980	0,8281
7	1,0000	1,0000	1,0000	0,9999	0,9996	0,9984	0,9952	0,9877	0,9726	0,9453
8	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9995	0,9983	0,9955	0,9893
9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9990
10	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

n=20

k \ p	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
0	0,3585	0,1216	0,0388	0,0115	0,0032	0,0008	0,0002	0,0000	0,0000	0,0000
1	0,7358	0,3917	0,1756	0,0692	0,0243	0,0076	0,0021	0,0005	0,0001	0,0000
2	0,9245	0,6769	0,4049	0,2061	0,0913	0,0355	0,0121	0,0036	0,0009	0,0002
3	0,9841	0,8670	0,6477	0,4114	0,2252	0,1071	0,0444	0,0160	0,0049	0,0013
4	0,9974	0,9568	0,8298	0,6296	0,4148	0,2375	0,1182	0,0510	0,0189	0,0059
5	0,9997	0,9887	0,9327	0,8042	0,6172	0,4164	0,2454	0,1256	0,0553	0,0207
6	1,0000	0,9976	0,9781	0,9133	0,7858	0,6080	0,4166	0,2500	0,1299	0,0577
7	1,0000	0,9996	0,9941	0,9679	0,8982	0,7723	0,6010	0,4159	0,2520	0,1316
8	1,0000	0,9999	0,9987	0,9900	0,9591	0,8867	0,7624	0,5956	0,4143	0,2517
9	1,0000	1,0000	0,9998	0,9974	0,9861	0,9520	0,8782	0,7553	0,5914	0,4119
10	1,0000	1,0000	1,0000	0,9994	0,9961	0,9829	0,9468	0,8725	0,7507	0,5881
11	1,0000	1,0000	1,0000	0,9999	0,9991	0,9949	0,9804	0,9435	0,8692	0,7483
12	1,0000	1,0000	1,0000	1,0000	0,9998	0,9987	0,9940	0,9790	0,9420	0,8684
13	1,0000	1,0000	1,0000	1,0000	1,0000	0,9997	0,9985	0,9935	0,9786	0,9423
14	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9997	0,9984	0,9936	0,9793
15	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9997	0,9985	0,9941
16	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9997	0,9987
17	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9998
18	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Table 2: Table de la loi binomiale (suite)

$$P(X=k)=C_n^k p^k (1-p)^{n-k}$$

(k le nombre d'occurrences parmi n)

n = 25

k	p 0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
0	0,2774	0,0718	0,0172	0,0038	0,0008	0,0001	0,0000	0,0000	0,0000	0,0000
1	0,6424	0,2712	0,0931	0,0274	0,0070	0,0016	0,0003	0,0001	0,0000	0,0000
2	0,8729	0,5371	0,2537	0,0982	0,0321	0,0090	0,0021	0,0004	0,0001	0,0000
3	0,9659	0,7636	0,4711	0,2340	0,0962	0,0332	0,0097	0,0024	0,0005	0,0001
4	0,9928	0,9020	0,6821	0,4207	0,2137	0,0905	0,0320	0,0095	0,0023	0,0005
5	0,9988	0,9666	0,8385	0,6167	0,3783	0,1935	0,0826	0,0294	0,0086	0,0020
6	0,9998	0,9905	0,9305	0,7800	0,5611	0,3407	0,1734	0,0736	0,0258	0,0073
7	1,0000	0,9977	0,9745	0,8909	0,7265	0,5118	0,3061	0,1536	0,0639	0,0216
8	1,0000	0,9995	0,9920	0,9532	0,8506	0,6769	0,4668	0,2735	0,1340	0,0539
9	1,0000	0,9999	0,9979	0,9827	0,9287	0,8106	0,6303	0,4246	0,2424	0,1148
10	1,0000	1,0000	0,9995	0,9944	0,9703	0,9022	0,7712	0,5858	0,3843	0,2122
11	1,0000	1,0000	0,9999	0,9985	0,9893	0,9558	0,8746	0,7323	0,5426	0,3450
12	1,0000	1,0000	1,0000	0,9996	0,9966	0,9825	0,9396	0,8462	0,6937	0,5000
13	1,0000	1,0000	1,0000	0,9999	0,9991	0,9940	0,9745	0,9222	0,8173	0,6550
14	1,0000	1,0000	1,0000	1,0000	0,9998	0,9982	0,9907	0,9656	0,9040	0,7878
15	1,0000	1,0000	1,0000	1,0000	1,0000	0,9995	0,9971	0,9868	0,9560	0,8852
16	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9992	0,9957	0,9826	0,9461
17	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9998	0,9988	0,9942	0,9784
18	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9997	0,9984	0,9927
19	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9996	0,9980
20	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9995
21	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999
22	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Table 3: Table de la loi binomiale (suite)

$$P(X = k) = C_n^k p^k (1-p)^{n-k}$$

(k le nombre d'occurrences parmi n)

n = 50

k \ p	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
0	0,0769	0,0052	0,0003	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
1	0,2794	0,0338	0,0029	0,0002	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
2	0,5405	0,1117	0,0142	0,0013	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000
3	0,7604	0,2503	0,0460	0,0057	0,0005	0,0000	0,0000	0,0000	0,0000	0,0000
4	0,8964	0,4312	0,1121	0,0185	0,0021	0,0002	0,0000	0,0000	0,0000	0,0000
5	0,9622	0,6161	0,2194	0,0480	0,0070	0,0007	0,0001	0,0000	0,0000	0,0000
6	0,9882	0,7702	0,3613	0,1034	0,0194	0,0025	0,0002	0,0000	0,0000	0,0000
7	0,9968	0,8779	0,5188	0,1904	0,0453	0,0073	0,0008	0,0001	0,0000	0,0000
8	0,9992	0,9421	0,6681	0,3073	0,0916	0,0183	0,0025	0,0002	0,0000	0,0000
9	0,9998	0,9755	0,7911	0,4437	0,1637	0,0402	0,0067	0,0008	0,0001	0,0000
10	1,0000	0,9906	0,8801	0,5836	0,2622	0,0789	0,0160	0,0022	0,0002	0,0000
11	1,0000	0,9968	0,9372	0,7107	0,3816	0,1390	0,0342	0,0057	0,0006	0,0000
12	1,0000	0,9990	0,9699	0,8139	0,5110	0,2229	0,0661	0,0133	0,0018	0,0002
13	1,0000	0,9997	0,9868	0,8894	0,6370	0,3279	0,1163	0,0280	0,0045	0,0005
14	1,0000	0,9999	0,9947	0,9393	0,7481	0,4468	0,1878	0,0540	0,0104	0,0013
15	1,0000	1,0000	0,9981	0,9692	0,8369	0,5692	0,2801	0,0955	0,0220	0,0033
16	1,0000	1,0000	0,9993	0,9856	0,9017	0,6839	0,3889	0,1561	0,0427	0,0077
17	1,0000	1,0000	0,9998	0,9937	0,9449	0,7822	0,5060	0,2369	0,0765	0,0164
18	1,0000	1,0000	0,9999	0,9975	0,9713	0,8594	0,6216	0,3356	0,1273	0,0325
19	1,0000	1,0000	1,0000	0,9991	0,9861	0,9152	0,7264	0,4465	0,1974	0,0595
20	1,0000	1,0000	1,0000	0,9997	0,9937	0,9522	0,8139	0,5610	0,2862	0,1013
21	1,0000	1,0000	1,0000	0,9999	0,9974	0,9749	0,8813	0,6701	0,3900	0,1611
22	1,0000	1,0000	1,0000	1,0000	0,9990	0,9877	0,9290	0,7660	0,5019	0,2399
23	1,0000	1,0000	1,0000	1,0000	0,9996	0,9944	0,9604	0,8438	0,6134	0,3359
24	1,0000	1,0000	1,0000	1,0000	0,9999	0,9976	0,9793	0,9022	0,7160	0,4439
25	1,0000	1,0000	1,0000	1,0000	1,0000	0,9991	0,9900	0,9427	0,8034	0,5561
26	1,0000	1,0000	1,0000	1,0000	1,0000	0,9997	0,9955	0,9686	0,8721	0,6641
27	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9981	0,9840	0,9220	0,7601
28	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9993	0,9924	0,9556	0,8389
29	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9997	0,9966	0,9765	0,8987
30	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9986	0,9884	0,9405
31	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9995	0,9947	0,9675
32	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9998	0,9978	0,9836
33	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9991	0,9923
34	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9997	0,9967
35	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9987
36	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9995
37	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9998
38	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Table 4: Table de la loi de Poisson

$$P(X = k) = e^{-\mu} \frac{\mu^k}{k!}$$

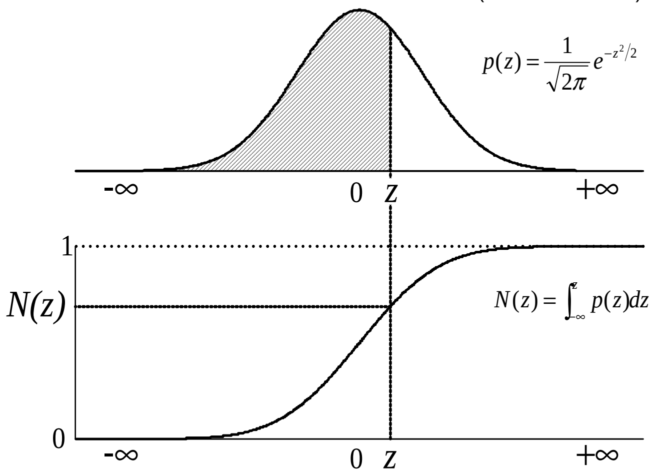
(μ le nombre d'occurrences moyen)

x \ μ	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066	0,3679
1	0,9953	0,9825	0,9631	0,9384	0,9098	0,8781	0,8442	0,8088	0,7725	0,7358
2	0,9998	0,9989	0,9964	0,9921	0,9856	0,9769	0,9659	0,9526	0,9371	0,9197
3	1,0000	0,9999	0,9997	0,9992	0,9982	0,9966	0,9942	0,9909	0,9865	0,9810
4	1,0000	1,0000	1,0000	0,9999	0,9998	0,9996	0,9992	0,9986	0,9977	0,9963
5	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9997	0,9994
6	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999
7	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

x \ μ	1,5	2	3	4	5	6	7	8	9	10
0	0,2231	0,1353	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001	0,0000
1	0,5578	0,4060	0,1991	0,0916	0,0404	0,0174	0,0073	0,0030	0,0012	0,0005
2	0,8088	0,6767	0,4232	0,2381	0,1247	0,0620	0,0296	0,0138	0,0062	0,0028
3	0,9344	0,8571	0,6472	0,4335	0,2650	0,1512	0,0818	0,0424	0,0212	0,0103
4	0,9814	0,9473	0,8153	0,6288	0,4405	0,2851	0,1730	0,0996	0,0550	0,0293
5	0,9955	0,9834	0,9161	0,7851	0,6160	0,4457	0,3007	0,1912	0,1157	0,0671
6	0,9991	0,9955	0,9665	0,8893	0,7622	0,6063	0,4497	0,3134	0,2068	0,1301
7	0,9998	0,9989	0,9881	0,9489	0,8666	0,7440	0,5987	0,4530	0,3239	0,2202
8	1,0000	0,9998	0,9962	0,9786	0,9319	0,8472	0,7291	0,5925	0,4557	0,3328
9	1,0000	1,0000	0,9989	0,9919	0,9682	0,9161	0,8305	0,7166	0,5874	0,4579
10	1,0000	1,0000	0,9997	0,9972	0,9863	0,9574	0,9015	0,8159	0,7060	0,5830
11	1,0000	1,0000	0,9999	0,9991	0,9945	0,9799	0,9467	0,8881	0,8030	0,6968
12	1,0000	1,0000	1,0000	0,9997	0,9980	0,9912	0,9730	0,9362	0,8758	0,7916
13	1,0000	1,0000	1,0000	0,9999	0,9993	0,9964	0,9872	0,9658	0,9261	0,8645
14	1,0000	1,0000	1,0000	1,0000	0,9998	0,9986	0,9943	0,9827	0,9585	0,9165
15	1,0000	1,0000	1,0000	1,0000	0,9999	0,9995	0,9976	0,9918	0,9780	0,9513
16	1,0000	1,0000	1,0000	1,0000	1,0000	0,9998	0,9990	0,9963	0,9889	0,9730
17	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9996	0,9984	0,9947	0,9857
18	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9993	0,9976	0,9928
19	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9997	0,9989	0,9965
20	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9996	0,9984
21	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9998	0,9993
22	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997
23	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999
24	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Table 5: Table de la loi normale centrée réduite

Probabilité qu'une variable aléatoire continue suivant une loi normale standard (ou centrée réduite) soit inférieure au seuil z.



La table ci-dessous présente les valeurs pour z positif. Pour z négatif la valeur est $N(z) = 1 - N(-z)$

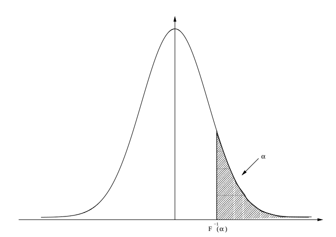
z	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0,0	0,500000	0,503989	0,507978	0,511966	0,515953	0,519939	0,523922	0,527903	0,531881	0,535856
0,1	0,539828	0,543795	0,547758	0,551717	0,555670	0,559618	0,563559	0,567495	0,571424	0,575345
0,2	0,579260	0,583166	0,587064	0,590954	0,594835	0,598706	0,602568	0,606420	0,610261	0,614092
0,3	0,617911	0,621720	0,625516	0,629300	0,633072	0,636831	0,640576	0,644309	0,648027	0,651732
0,4	0,655422	0,659097	0,662757	0,666402	0,670031	0,673645	0,677242	0,680822	0,684386	0,687933
0,5	0,691462	0,694974	0,698468	0,701944	0,705401	0,708840	0,712260	0,715661	0,719043	0,722405
0,6	0,725747	0,729069	0,732371	0,735653	0,738914	0,742154	0,745373	0,748571	0,751748	0,754903
0,7	0,758036	0,761148	0,764238	0,767305	0,770350	0,773373	0,776373	0,779350	0,782305	0,785236
0,8	0,788145	0,791030	0,793892	0,796731	0,799546	0,802337	0,805105	0,807850	0,810570	0,813267
0,9	0,815940	0,818589	0,821214	0,823814	0,826391	0,828944	0,831472	0,833977	0,836457	0,838913
1,0	0,841345	0,843752	0,846136	0,848495	0,850830	0,853141	0,855428	0,857690	0,859929	0,862143
1,1	0,864334	0,866500	0,868643	0,870762	0,872857	0,874928	0,876976	0,879000	0,881000	0,882977
1,2	0,884930	0,886861	0,888768	0,890651	0,892512	0,894350	0,896165	0,897958	0,899727	0,901475
1,3	0,903200	0,904902	0,906582	0,908241	0,909877	0,911492	0,913085	0,914657	0,916207	0,917736
1,4	0,919243	0,920730	0,922196	0,923641	0,925066	0,926471	0,927855	0,929219	0,930563	0,931888
1,5	0,933193	0,934478	0,935745	0,936992	0,938220	0,939429	0,940620	0,941792	0,942947	0,944083
1,6	0,945201	0,946301	0,947384	0,948449	0,949497	0,950529	0,951543	0,952540	0,953521	0,954486
1,7	0,955435	0,956367	0,957284	0,958185	0,959070	0,959941	0,960796	0,961636	0,962462	0,963273
1,8	0,964070	0,964852	0,965620	0,966375	0,967116	0,967843	0,968557	0,969258	0,969946	0,970621
1,9	0,971283	0,971933	0,972571	0,973197	0,973810	0,974412	0,975002	0,975581	0,976148	0,976705
2,0	0,977250	0,977784	0,978308	0,978822	0,979325	0,979818	0,980301	0,980774	0,981237	0,981691
2,1	0,982136	0,982571	0,982997	0,983414	0,983823	0,984222	0,984614	0,984997	0,985371	0,985738
2,2	0,986097	0,986447	0,986791	0,987126	0,987455	0,987776	0,988089	0,988396	0,988696	0,988989
2,3	0,989276	0,989556	0,989830	0,990097	0,990358	0,990613	0,990863	0,991106	0,991344	0,991576
2,4	0,991802	0,992024	0,992240	0,992451	0,992656	0,992857	0,993053	0,993244	0,993431	0,993613
2,5	0,993790	0,993963	0,994132	0,994297	0,994457	0,994614	0,994766	0,994915	0,995060	0,995201
2,6	0,995339	0,995473	0,995604	0,995731	0,995855	0,995975	0,996093	0,996207	0,996319	0,996427
2,7	0,996533	0,996636	0,996736	0,996833	0,996928	0,997020	0,997110	0,997197	0,997282	0,997365
2,8	0,997445	0,997523	0,997599	0,997673	0,997744	0,997814	0,997882	0,997948	0,998012	0,998074
2,9	0,998134	0,998193	0,998250	0,998305	0,998359	0,998411	0,998462	0,998511	0,998559	0,998605

Pour les valeurs de z supérieures à 3 :

z	**3,0**	**3,1**	**3,2**	**3,3**	**3,4**	**3,5**	**3,6**	**3,8**	**4,0**	**4,5**
N(z)	0,998650	0,999032	0,999313	0,999517	0,999663	0,999767	0,999841	0,999928	0,999968	0,999997

Table 6: Table de la loi normale : quantiles

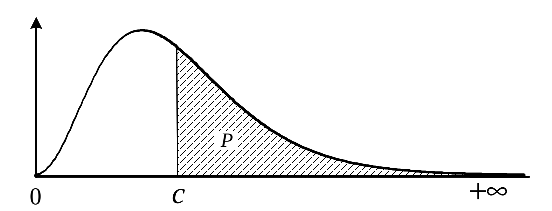
Pour une valeur $\alpha \in]0; 0.5[$, la table ci-dessous renvoie la valeur $F^{-1}(\alpha)$ de la fonction quantile F^{-1} de la loi normale centrée réduite au point α.



α	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
0.00	∞	3.0902	2.8782	2.7478	2.6521	2.5758	2.5121	2.4573	2.4089	2.3656
0.01	2.3263	2.2904	2.2571	2.2262	2.1973	2.1701	2.1444	2.1201	2.0969	2.0749
0.02	2.0537	2.0335	2.0141	1.9954	1.9774	1.9600	1.9431	1.9268	1.9110	1.8957
0.03	1.8808	1.8663	1.8522	1.8384	1.8250	1.8119	1.7991	1.7866	1.7744	1.7624
0.04	1.7507	1.7392	1.7279	1.7169	1.7060	1.6954	1.6849	1.6747	1.6646	1.6546
0.05	1.6449	1.6352	1.6258	1.6164	1.6072	1.5982	1.5893	1.5805	1.5718	1.5632
0.06	1.5548	1.5464	1.5382	1.5301	1.5220	1.5141	1.5063	1.4985	1.4909	1.4833
0.07	1.4758	1.4684	1.4611	1.4538	1.4466	1.4395	1.4325	1.4255	1.4187	1.4118
0.08	1.4051	1.3984	1.3917	1.3852	1.3787	1.3722	1.3658	1.3595	1.3532	1.3469
0.09	1.3408	1.3346	1.3285	1.3225	1.3165	1.3106	1.3047	1.2988	1.2930	1.2873
0.10	1.2816	1.2759	1.2702	1.2646	1.2591	1.2536	1.2481	1.2426	1.2372	1.2319
0.11	1.2265	1.2212	1.2160	1.2107	1.2055	1.2004	1.1952	1.1901	1.1850	1.1800
0.12	1.1750	1.1700	1.1650	1.1601	1.1552	1.1503	1.1455	1.1407	1.1359	1.1311
0.13	1.1264	1.1217	1.1170	1.1123	1.1077	1.1031	1.0985	1.0939	1.0893	1.0848
0.14	1.0803	1.0758	1.0714	1.0669	1.0625	1.0581	1.0537	1.0494	1.0450	1.0407
0.15	1.0364	1.0322	1.0279	1.0237	1.0194	1.0152	1.0110	1.0069	1.0027	0.9986
0.16	0.9945	0.9904	0.9863	0.9822	0.9782	0.9741	0.9701	0.9661	0.9621	0.9581
0.17	0.9542	0.9502	0.9463	0.9424	0.9385	0.9346	0.9307	0.9269	0.9230	0.9192
0.18	0.9154	0.9116	0.9078	0.9040	0.9002	0.8965	0.8927	0.8890	0.8853	0.8816
0.19	0.8779	0.8742	0.8705	0.8669	0.8633	0.8596	0.8560	0.8524	0.8488	0.8452
0.20	0.8416	0.8381	0.8345	0.8310	0.8274	0.8239	0.8204	0.8169	0.8134	0.8099
0.21	0.8064	0.8030	0.7995	0.7961	0.7926	0.7892	0.7858	0.7824	0.7790	0.7756
0.22	0.7722	0.7688	0.7655	0.7621	0.7588	0.7554	0.7521	0.7488	0.7454	0.7421
0.23	0.7388	0.7356	0.7323	0.7290	0.7257	0.7225	0.7192	0.7160	0.7128	0.7095
0.24	0.7063	0.7031	0.6999	0.6967	0.6935	0.6903	0.6871	0.6840	0.6808	0.6776
0.25	0.6745	0.6713	0.6682	0.6651	0.6620	0.6588	0.6557	0.6526	0.6495	0.6464
0.26	0.6433	0.6403	0.6372	0.6341	0.6311	0.6280	0.6250	0.6219	0.6189	0.6158
0.27	0.6128	0.6098	0.6068	0.6038	0.6008	0.5978	0.5948	0.5918	0.5888	0.5858
0.28	0.5828	0.5799	0.5769	0.5740	0.5710	0.5681	0.5651	0.5622	0.5592	0.5563
0.29	0.5534	0.5505	0.5476	0.5446	0.5417	0.5388	0.5359	0.5330	0.5302	0.5273
0.30	0.5244	0.5215	0.5187	0.5158	0.5129	0.5101	0.5072	0.5044	0.5015	0.4987
0.31	0.4959	0.4930	0.4902	0.4874	0.4845	0.4817	0.4789	0.4761	0.4733	0.4705
0.32	0.4677	0.4649	0.4621	0.4593	0.4565	0.4538	0.4510	0.4482	0.4454	0.4427
0.33	0.4399	0.4372	0.4344	0.4316	0.4289	0.4261	0.4234	0.4207	0.4179	0.4152
0.34	0.4125	0.4097	0.4070	0.4043	0.4016	0.3989	0.3961	0.3934	0.3907	0.3880
0.35	0.3853	0.3826	0.3799	0.3772	0.3745	0.3719	0.3692	0.3665	0.3638	0.3611
0.36	0.3585	0.3558	0.3531	0.3505	0.3478	0.3451	0.3425	0.3398	0.3372	0.3345
0.37	0.3319	0.3292	0.3266	0.3239	0.3213	0.3186	0.3160	0.3134	0.3107	0.3081
0.38	0.3055	0.3029	0.3002	0.2976	0.2950	0.2924	0.2898	0.2871	0.2845	0.2819
0.39	0.2793	0.2767	0.2741	0.2715	0.2689	0.2663	0.2637	0.2611	0.2585	0.2559
0.40	0.2533	0.2508	0.2482	0.2456	0.2430	0.2404	0.2378	0.2353	0.2327	0.2301
0.41	0.2275	0.2250	0.2224	0.2198	0.2173	0.2147	0.2121	0.2096	0.2070	0.2045
0.42	0.2019	0.1993	0.1968	0.1942	0.1917	0.1891	0.1866	0.1840	0.1815	0.1789
0.43	0.1764	0.1738	0.1713	0.1687	0.1662	0.1637	0.1611	0.1586	0.1560	0.1535
0.44	0.1510	0.1484	0.1459	0.1434	0.1408	0.1383	0.1358	0.1332	0.1307	0.1282
0.45	0.1257	0.1231	0.1206	0.1181	0.1156	0.1130	0.1105	0.1080	0.1055	0.1030
0.46	0.1004	0.0979	0.0954	0.0929	0.0904	0.0878	0.0853	0.0828	0.0803	0.0778
0.47	0.0753	0.0728	0.0702	0.0677	0.0652	0.0627	0.0602	0.0577	0.0552	0.0527
0.48	0.0502	0.0476	0.0451	0.0426	0.0401	0.0376	0.0351	0.0326	0.0301	0.0276
0.49	0.0251	0.0226	0.0201	0.0175	0.0150	0.0125	0.0100	0.0075	0.0050	0.0025

Table 7: Table de la loi du χ^2

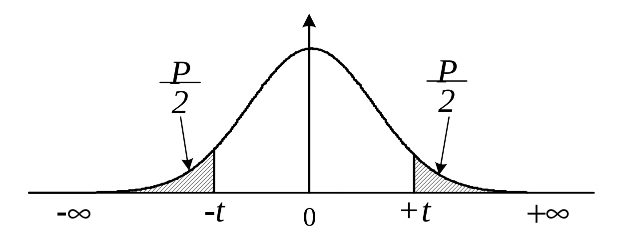
Valeurs de c ayant la probabilité P d'être dépassées en valeur absolue.



r	P= 0,990	P= 0,975	P= 0,950	P= 0,900	P= 0,800	P= 0,700	P= 0,500	P= 0,300	P= 0,200	P= 0,100	P= 0,010	P= 0,005	P= 0,001
1	0,000	0,001	0,004	0,016	0,064	0,148	0,455	1,074	1,642	2,706	6,635	7,879	10,828
2	0,020	0,051	0,103	0,211	0,446	0,713	1,386	2,408	3,219	4,605	9,210	10,597	13,816
3	0,115	0,216	0,352	0,584	1,005	1,424	2,366	3,665	4,642	6,251	11,345	12,838	16,266
4	0,297	0,484	0,711	1,064	1,649	2,195	3,357	4,878	5,989	7,779	13,277	14,860	18,467
5	0,554	0,831	1,145	1,610	2,343	3,000	4,351	6,064	7,289	9,236	15,086	16,750	20,515
6	0,872	1,237	1,635	2,204	3,070	3,828	5,348	7,231	8,558	10,645	16,812	18,548	22,458
7	1,239	1,690	2,167	2,833	3,822	4,671	6,346	8,383	9,803	12,017	18,475	20,278	24,322
8	1,646	2,180	2,733	3,490	4,594	5,527	7,344	9,524	11,030	13,362	20,090	21,955	26,124
9	2,088	2,700	3,325	4,168	5,380	6,393	8,343	10,656	12,242	14,684	21,666	23,589	27,877
10	2,558	3,247	3,940	4,865	6,179	7,267	9,342	11,781	13,442	15,987	23,209	25,188	29,588
11	3,053	3,816	4,575	5,578	6,989	8,148	10,341	12,899	14,631	17,275	24,725	26,757	31,264
12	3,571	4,404	5,226	6,304	7,807	9,034	11,340	14,011	15,812	18,549	26,217	28,300	32,909
13	4,107	5,009	5,892	7,042	8,634	9,926	12,340	15,119	16,985	19,812	27,688	29,819	34,528
14	4,660	5,629	6,571	7,790	9,467	10,821	13,339	16,222	18,151	21,064	29,141	31,319	36,123
15	5,229	6,262	7,261	8,547	10,307	11,721	14,339	17,322	19,311	22,307	30,578	32,801	37,697
16	5,812	6,908	7,962	9,312	11,152	12,624	15,338	18,418	20,465	23,542	32,000	34,267	39,252
17	6,408	7,564	8,672	10,085	12,002	13,531	16,338	19,511	21,615	24,769	33,409	35,718	40,790
18	7,015	8,231	9,390	10,865	12,857	14,440	17,338	20,601	22,760	25,989	34,805	37,156	42,312
19	7,633	8,907	10,117	11,651	13,716	15,352	18,338	21,689	23,900	27,204	36,191	38,582	43,820
20	8,260	9,591	10,851	12,443	14,578	16,266	19,337	22,775	25,038	28,412	37,566	39,997	45,315
21	8,897	10,283	11,591	13,240	15,445	17,182	20,337	23,858	26,171	29,615	38,932	41,401	46,797
22	9,542	10,982	12,338	14,041	16,314	18,101	21,337	24,939	27,301	30,813	40,289	42,796	48,268
23	10,196	11,689	13,091	14,848	17,187	19,021	22,337	26,018	28,429	32,007	41,638	44,181	49,728
24	10,856	12,401	13,848	15,659	18,062	19,943	23,337	27,096	29,553	33,196	42,980	45,559	51,179
25	11,524	13,120	14,611	16,473	18,940	20,867	24,337	28,172	30,675	34,382	44,314	46,928	52,620
26	12,198	13,844	15,379	17,292	19,820	21,792	25,336	29,246	31,795	35,563	45,642	48,290	54,052
27	12,879	14,573	16,151	18,114	20,703	22,719	26,336	30,319	32,912	36,741	46,963	49,645	55,476
28	13,565	15,308	16,928	18,939	21,588	23,647	27,336	31,391	34,027	37,916	48,278	50,993	56,892
29	14,256	16,047	17,708	19,768	22,475	24,577	28,336	32,461	35,139	39,087	49,588	52,336	58,301
30	14,953	16,791	18,493	20,599	23,364	25,508	29,336	33,530	36,250	40,256	50,892	53,672	59,703
40	22,164	24,433	26,509	29,051	32,345	34,872	39,335	44,165	47,269	51,805	63,691	66,766	73,402
80	53,540	57,153	60,391	64,278	69,207	72,915	79,334	86,120	90,405	96,578	112,329	116,321	124,839
120	86,923	91,573	95,705	100,624	106,806	111,419	119,334	127,616	132,806	140,233	158,950	163,648	173,617

Table 8: Table de la loi de Student

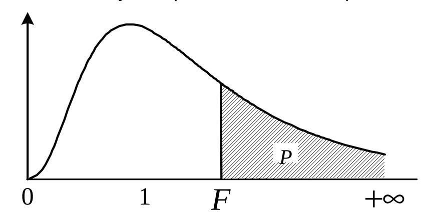
Valeurs de t ayant la probabilité P d'être dépassées en valeur absolue.



r	P= 0,90	P= 0,80	P= 0,70	P= 0,60	P= 0,50	P= 0,40	P= 0,30	P= 0,20	P= 0,10	P= 0,05	P= 0,01	P= 0,005
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	63,657	127,321
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	9,925	14,089
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	5,841	7,453
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	4,604	5,598
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	4,032	4,773
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,707	4,317
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	3,499	4,029
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	3,355	3,833
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	3,250	3,690
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	3,169	3,581
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	3,106	3,497
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	3,055	3,428
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	3,012	3,372
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,977	3,326
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,947	3,286
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,921	3,252
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,898	3,222
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,878	3,197
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,861	3,174
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,845	3,153
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,831	3,135
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,819	3,119
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,807	3,104
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,797	3,091
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,787	3,078
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,779	3,067
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,771	3,057
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,763	3,047
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,756	3,038
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,750	3,030
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,704	2,971
80	0,126	0,254	0,387	0,526	0,678	0,846	1,043	1,292	1,664	1,990	2,639	2,887
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,617	2,860
∞	0,126	0,253	0,385	0,524	0,675	0,842	1,036	1,282	1,645	1,960	2,576	2,808

Table 9: Table de la loi de Fisher-Snedecor

Valeurs de F ayant la probabilité P d'être dépassées en valeur absolue.



ν_2	$\nu_1 = 1$		$\nu_1 = 2$		$\nu_1 = 3$		$\nu_1 = 4$		$\nu_1 = 5$		$\nu_1 = 6$	
	P= 0,05	**P= 0,01**	**P= 0,05**	**P= 0,01**	**P= 0,05**	**P= 0,01**	**P= 0,05**	**P= 0,01**	**P= 0,05**	**P= 0,01**	**P= 0,05**	**P= 0,01**
1	161,448	4052,181	199,500	4999,500	215,707	5403,352	224,583	5624,583	230,162	5763,650	233,986	5858,986
2	18,513	98,503	19,000	99,000	19,164	99,166	19,247	99,249	19,296	99,299	19,330	99,333
3	10,128	34,116	9,552	30,817	9,277	29,457	9,117	28,710	9,013	28,237	8,941	27,911
4	7,709	21,198	6,944	18,000	6,591	16,694	6,388	15,977	6,256	15,522	6,163	15,207
5	6,608	16,258	5,786	13,274	5,409	12,060	5,192	11,392	5,050	10,967	4,950	10,672
6	5,987	13,745	5,143	10,925	4,757	9,780	4,534	9,148	4,387	8,746	4,284	8,466
7	5,591	12,246	4,737	9,547	4,347	8,451	4,120	7,847	3,972	7,460	3,866	7,191
8	5,318	11,259	4,459	8,649	4,066	7,591	3,838	7,006	3,687	6,632	3,581	6,371
9	5,117	10,561	4,256	8,022	3,863	6,992	3,633	6,422	3,482	6,057	3,374	5,802
10	4,965	10,044	4,103	7,559	3,708	6,552	3,478	5,994	3,326	5,636	3,217	5,386
11	4,844	9,646	3,982	7,206	3,587	6,217	3,357	5,668	3,204	5,316	3,095	5,069
12	4,747	9,330	3,885	6,927	3,490	5,953	3,259	5,412	3,106	5,064	2,996	4,821
13	4,667	9,074	3,806	6,701	3,411	5,739	3,179	5,205	3,025	4,862	2,915	4,620
14	4,600	8,862	3,739	6,515	3,344	5,564	3,112	5,035	2,958	4,695	2,848	4,456
15	4,543	8,683	3,682	6,359	3,287	5,417	3,056	4,893	2,901	4,556	2,790	4,318
16	4,494	8,531	3,634	6,226	3,239	5,292	3,007	4,773	2,852	4,437	2,741	4,202
17	4,451	8,400	3,592	6,112	3,197	5,185	2,965	4,669	2,810	4,336	2,699	4,102
18	4,414	8,285	3,555	6,013	3,160	5,092	2,928	4,579	2,773	4,248	2,661	4,015
19	4,381	8,185	3,522	5,926	3,127	5,010	2,895	4,500	2,740	4,171	2,628	3,939
20	4,351	8,096	3,493	5,849	3,098	4,938	2,866	4,431	2,711	4,103	2,599	3,871
21	4,325	8,017	3,467	5,780	3,072	4,874	2,840	4,369	2,685	4,042	2,573	3,812
22	4,301	7,945	3,443	5,719	3,049	4,817	2,817	4,313	2,661	3,988	2,549	3,758
23	4,279	7,881	3,422	5,664	3,028	4,765	2,796	4,264	2,640	3,939	2,528	3,710
24	4,260	7,823	3,403	5,614	3,009	4,718	2,776	4,218	2,621	3,895	2,508	3,667
25	4,242	7,770	3,385	5,568	2,991	4,675	2,759	4,177	2,603	3,855	2,490	3,627
26	4,225	7,721	3,369	5,526	2,975	4,637	2,743	4,140	2,587	3,818	2,474	3,591
27	4,210	7,677	3,354	5,488	2,960	4,601	2,728	4,106	2,572	3,785	2,459	3,558
28	4,196	7,636	3,340	5,453	2,947	4,568	2,714	4,074	2,558	3,754	2,445	3,528
29	4,183	7,598	3,328	5,420	2,934	4,538	2,701	4,045	2,545	3,725	2,432	3,499
30	4,171	7,562	3,316	5,390	2,922	4,510	2,690	4,018	2,534	3,699	2,421	3,473
40	4,085	7,314	3,232	5,179	2,839	4,313	2,606	3,828	2,449	3,514	2,336	3,291
80	3,960	6,963	3,111	4,881	2,719	4,036	2,486	3,563	2,329	3,255	2,214	3,036
120	3,920	6,851	3,072	4,787	2,680	3,949	2,447	3,480	2,290	3,174	2,175	2,956
∞	3,842	6,637	2,997	4,607	2,606	3,784	2,373	3,321	2,215	3,019	2,099	2,804

ν_2	$\nu_1 = 8$		$\nu_1 = 10$		$\nu_1 = 12$		$\nu_1 = 24$		$\nu_1 = 48$		$\nu_1 = \infty$	
	P= 0,05	**P= 0,01**	**P= 0,05**	**P= 0,01**	**P= 0,05**	**P= 0,01**	**P= 0,05**	**P= 0,01**	**P= 0,05**	**P= 0,01**	**P= 0,05**	**P= 0,01**
1	238,883	5981,070	241,882	6055,847	243,906	6106,321	249,052	6234,631	251,669	6299,892	254,314	6365,861
2	19,371	99,374	19,396	99,399	19,413	99,416	19,454	99,458	19,475	99,478	19,496	99,499
3	8,845	27,489	8,786	27,229	8,745	27,052	8,639	26,598	8,583	26,364	8,526	26,125
4	6,041	14,799	5,964	14,546	5,912	14,374	5,774	13,929	5,702	13,699	5,628	13,463
5	4,818	10,289	4,735	10,051	4,678	9,888	4,527	9,466	4,448	9,247	4,365	9,020
6	4,147	8,102	4,060	7,874	4,000	7,718	3,841	7,313	3,757	7,100	3,669	6,880
7	3,726	6,840	3,637	6,620	3,575	6,469	3,410	6,074	3,322	5,866	3,230	5,650
8	3,438	6,029	3,347	5,814	3,284	5,667	3,115	5,279	3,024	5,074	2,928	4,859
9	3,230	5,467	3,137	5,257	3,073	5,111	2,900	4,729	2,807	4,525	2,707	4,311
10	3,072	5,057	2,978	4,849	2,913	4,706	2,737	4,327	2,641	4,124	2,538	3,909
11	2,948	4,744	2,854	4,539	2,788	4,397	2,609	4,021	2,511	3,818	2,404	3,602
12	2,849	4,499	2,753	4,296	2,687	4,155	2,505	3,780	2,405	3,578	2,296	3,361
13	2,767	4,302	2,671	4,100	2,604	3,960	2,420	3,587	2,318	3,384	2,206	3,165
14	2,699	4,140	2,602	3,939	2,534	3,800	2,349	3,427	2,245	3,224	2,131	3,004
15	2,641	4,004	2,544	3,805	2,475	3,666	2,288	3,294	2,182	3,090	2,066	2,868
16	2,591	3,890	2,494	3,691	2,425	3,553	2,235	3,181	2,128	2,976	2,010	2,753
17	2,548	3,791	2,450	3,593	2,381	3,455	2,190	3,084	2,081	2,878	1,960	2,653
18	2,510	3,705	2,412	3,508	2,342	3,371	2,150	2,999	2,040	2,793	1,917	2,566
19	2,477	3,631	2,378	3,434	2,308	3,297	2,114	2,925	2,003	2,718	1,878	2,489
20	2,447	3,564	2,348	3,368	2,278	3,231	2,082	2,859	1,970	2,652	1,843	2,421
21	2,420	3,506	2,321	3,310	2,250	3,173	2,054	2,801	1,941	2,593	1,812	2,360
22	2,397	3,453	2,297	3,258	2,226	3,121	2,028	2,749	1,914	2,540	1,783	2,305
23	2,375	3,406	2,275	3,211	2,204	3,074	2,005	2,702	1,890	2,492	1,757	2,256
24	2,355	3,363	2,255	3,168	2,183	3,032	1,984	2,659	1,868	2,448	1,733	2,211
25	2,337	3,324	2,236	3,129	2,165	2,993	1,964	2,620	1,847	2,409	1,711	2,169
26	2,321	3,288	2,220	3,094	2,148	2,958	1,946	2,585	1,828	2,373	1,691	2,131
27	2,305	3,256	2,204	3,062	2,132	2,926	1,930	2,552	1,811	2,339	1,672	2,097
28	2,291	3,226	2,190	3,032	2,118	2,896	1,915	2,522	1,795	2,309	1,654	2,064
29	2,278	3,198	2,177	3,005	2,104	2,868	1,901	2,495	1,780	2,280	1,638	2,034
30	2,266	3,173	2,165	2,979	2,092	2,843	1,887	2,469	1,766	2,254	1,622	2,006
40	2,180	2,993	2,077	2,801	2,003	2,665	1,793	2,288	1,666	2,068	1,509	1,805
80	2,056	2,742	1,951	2,551	1,875	2,415	1,654	2,032	1,514	1,799	1,325	1,494
120	2,016	2,663	1,910	2,472	1,834	2,336	1,608	1,950	1,463	1,711	1,254	1,381
∞	1,939	2,513	1,832	2,323	1,753	2,187	1,518	1,793	1,359	1,537	1,000	1,000

POUR ALLER PLUS LOIN

La méthode de Holt-Winters

Les équations (4.15) et (4.16) du LED peuvent encore s'écrire sous la forme suivante[12] :

$$\hat{d}_t = \theta Y_t + (1-\theta)(\hat{d}_{t-1} + \hat{a}_{t-1}) = \theta Y_t + (1-\theta)\hat{Y}_{t-1} \quad (4.36)$$

et

$$\hat{a}_t = \lambda(\hat{d}_t - \hat{d}_{t-1}) + (1-\lambda)\hat{a}_{t-1} \quad (4.37)$$

avec $\theta = \alpha(2-\alpha)$ et $\lambda = \frac{\alpha}{2-\alpha}$; $0 < \theta < 1$ et $0 < \lambda < 1$. La **méthode de Holt** consiste à utiliser ces deux dernières équations du LED pour estimer la tendance, mais en choisissant les deux constantes θ et λ indépendamment. Plus spécifiquement, l'équation (4.36) s'interprète comme une moyenne pondérée de deux informations relatives au niveau de la série à la date t et θ représente la pondération que l'on souhaite attribuer à la dernière valeur observée Y_t comparativement à la prévision réalisée à la date $t-1$. L'équation (4.37) s'interprète comme une moyenne pondérée de deux informations relatives à la pente de la série à la date t et λ renvoie au poids attribué à la dernière variation de la tendance $(\hat{d}_t - \hat{d}_{t-1})$ comparativement à l'estimation de la pente en $(t-1)$, $\hat{a}_{t-1}$. Le lissage de Holt comprend donc deux paramètres, un pour l'ordonnée à l'origine (c'est-à-dire le niveau de la série), l'autre pour la pente. La méthode de Holt consiste ainsi en une adaptation du LED visant à mieux prendre en compte les variations de la pente au cours du temps : plus la valeur retenue pour λ est élevée, plus la réaction à un changement de la pente de la tendance est rapide. Pour finir, notons que cette méthode présente l'avantage, par rapport au LED, de pouvoir être étendue à une série présentant une composante saisonnière ; un troisième paramètre, relatif à la saisonnalité, est alors pris en compte en plus des deux précédents paramètres (ordonnée à l'origine et pente). Cela conduit à la **méthode de Holt et Winters** (voir Winters, 1960) consistant en l'application d'un lissage de Holt sur la partie non saisonnière, couplée à un lissage exponentiel dit de Winters pour la partie saisonnière.

12 Pour la démonstration, voir Gouriéroux et Monfort (1995).

Trois questions à Axelle Chauvet-Peyrard, chef de division à l'INSEE et précédemment responsable de la méthodologie de l'indice des prix à la consommation au sein de la direction des statistiques démographiques et sociales de l'INSEE

1) Que mesure précisément l'indice des prix à la consommation mis à disposition par l'INSEE et comment est-il calculé en pratique ? Quels problèmes peut poser un tel indice agrégé ?

L'indice des prix à la consommation (IPC) est un indicateur synthétique dont l'objectif est d'estimer la pression inflationniste qui s'exerce à l'intérieur du territoire français, et ce à travers la mesure de l'évolution des prix des biens et services offerts à la consommation finale par les points de vente installés sur le territoire français (France métropolitaine et départements d'outre-mer). Le recensement complet de tous les produits offerts aux consommateurs étant matériellement impossible, cette mesure s'effectue par le biais d'un échantillonnage. Un « panier-type » de biens et services est ainsi défini et révisé chaque année de manière à être représentatif de la consommation réelle des ménages, telle qu'observée sur le territoire dans un passé récent[1]. Parallèlement, on effectue un échantillonnage des points de vente au sein desquels seront observés tous les mois les prix des produits retenus dans le panier-type. Ces observations de prix sont ensuite agrégées sous forme d'indice. La formule retenue pour l'IPC comme pour l'IPCH (indice des prix à la consommation « harmonisé » au niveau européen) est, conformément à la réglementation européenne, celle de Laspeyres. Cette formule, qui consiste à agréger les indices élémentaires de prix $I^i_{t/0}$ de différents produits *i* en effectuant une moyenne arithmétique pondérée (voir équation (3.22)), appelle plusieurs commentaires :

- L'indice compare les prix de la période courante (notée *t*) aux prix d'une période de référence (notée 0). Dans l'IPC, la période de référence est la même pendant toute l'année *n*, et correspond au mois de décembre de l'année *n-1*. L'IPC diffusé est en fait un indice « chaîné » de manière à présenter une référence temporelle unique[2] et donc à faciliter les comparaisons d'une année sur l'autre et les analyses sur longue période.
- Les pondérations α^i_0 utilisées dans le calcul de l'IPC correspondent aux parts de dépense consacrées en période de référence au produit i. Elles sont issues des comptes nationaux de l'année *n-2* et « valorisées » en décembre *n-1*. Ces données n'étant disponibles qu'à partir d'un certain niveau d'agrégation, les indices « élémentaires » $I^i_{t/0}$ utilisés dans le calcul de l'IPC sont déjà le résultat d'une agrégation, que l'on appelle agrégation élémentaire ou de premier niveau et qui, faute de données fines de pondérations, ne peut pas se faire selon la formule de Laspeyres. En pratique, on opte pour une formule de Dutot ou une formule de Jevons en fonction des produits[3].

Comme tout produit d'une agrégation, l'IPC reflète des réalités et des situations très diverses. Étant un indicateur d'inflation, il couvre l'intégralité du champ de la consommation des ménages, y compris des biens durables tels que les ordinateurs ou des services plus ou moins fréquemment consommés. Il y a souvent une mécompréhension de l'IPC, le grand public étant davantage attentif à ce qu'on appelle parfois le « prix du caddie ». De plus, pour être une

[1]On dispose pour ce faire des données issues des comptes de la nation ainsi que de sources complémentaires comme l'enquête quinquennale « budget des familles » par exemple.

[2]Actuellement, la période de référence prise pour la diffusion de l'IPC est l'année 1998. A partir du 1er janvier 2016, cette référence sera modifiée et deviendra l'année 2015.

[3]La littérature sur les indices de prix a largement étudié le problème de l'agrégation de premier niveau et comparé les qualités et défauts des différentes formules à disposition. La formule de Dutot consiste dans le rapport des moyennes arithmétiques simples des prix aux périodes courante et de référence. La formule de Jevons reprend cette formule, en remplaçant les moyennes arithmétiques par des moyennes géométriques.

mesure efficace de la pression inflationniste, il est nécessaire que la mesure d'évolution des prix se fasse *à qualité constante*. De cette façon, si le prix étiqueté d'un produit reste stable mais que sa qualité s'améliore, l'indice élémentaire du produit diminuera de fait. C'est ce qui se passe par exemple sur les ordinateurs. Et c'est également ce qui donne un IPC relativement stable depuis des années autour de 1,5 à 2 % d'inflation annuelle, alors que les ménages « vivent » une hausse des prix étiquetés jugée bien supérieure.

2) Pourquoi l'INSEE calcule-t-il également des indices des prix à la consommation par catégorie de ménage en tenant compte notamment de la région d'habitation et de la taille de la commune de référence ? Concrètement, que met en évidence le calcul de ces indices ?

Comme dit précédemment, l'IPC reflète des réalités diverses et potentiellement contrastées. Les pondérations prises pour le calcul de l'indice, et dont les valeurs ont un fort impact sur le résultat final, correspondent aux parts de budget consacrées à chaque type de produit, selon une nomenclature de produits « comptable » et considérées de manière agrégée pour l'ensemble de la population. Or les profils de consommation des Français sont en réalité très diversifiés. C'est pourquoi l'INSEE a par exemple mis à disposition sur son site internet www.insee.fr un simulateur d'indice personnalisé permettant de recalculer l'indice de Laspeyres en y imputant les parts de dépense réellement accordées par l'individu à tel ou tel poste de consommation. Ainsi, par exemple, l'évolution des loyers comptera pour zéro dans l'indice personnalisé d'un ménage propriétaire de son logement. Le mode de vie influence obligatoirement le mode de consommation. Selon la tranche d'âge, le milieu d'habitation (rural ou urbain) ou encore la composition du ménage (célibataires, couples avec ou sans enfants, familles monoparentales, etc.), les parts de dépense consacrées à tel ou tel type de produit varient. On peut donc être amené à s'interroger sur les différences éventuelles d'inflation subie par différentes catégories de ménages, afin de s'assurer qu'il n'y a pas de trop grandes inégalités au sein de la population française par rapport à l'inflation. Les indices par catégories de ménages, qui sont publiés annuellement par l'INSEE, montrent en effet que les disparités restent en général assez faibles. La région d'habitation, par exemple, joue assez peu sur l'inflation subie par les ménages, même si on constate que la facture énergétique pèse plus durement sur les ménages du nord et de l'est de la France. Plus notable est la différence d'inflation subie par les ménages en fonction de la taille de leur commune de résidence. Ce résultat semble lié au poids budgétaire que représentent les carburants ainsi que les combustibles liquides (fioul) dans le budget des ménages ruraux, qui est plus important que pour les ménages urbains. De ce fait, la forte inflation constatée ces dernières années sur ces deux postes a un impact plus fort sur le budget des ménages ruraux que sur celui des ménages urbains[4].

3) Au niveau méthodologique, pourquoi privilégier l'indice de Laspeyres pour le calcul des indices des prix à la consommation ?

L'indice de Laspeyres est un des plus anciens indices qui existe et le seul reconnu par la réglementation européenne. Par rapport à ses prédécesseurs, il possède le grand avantage d'être pondéré par les parts de dépense réellement consacrées à tel ou tel produit. Or, d'un point de vue statistique, ne pas pondérer les produits revient à supposer l'existence d'un modèle économétrique sous-jacent dans lequel les évolutions de prix des produits oscillent

---

[4] Sur ces questions relatives à l'impact de la région d'habitation et de la taille de la commune de résidence, on pourra consulter Chauvet-Peyrard (2013), « Indices des prix à la consommation 1998-2012 selon la région d'habitation des ménages et selon la taille de la commune de résidence », Document de travail INSEE, Direction des statistiques démographiques et sociales F1306.

autour d'un taux d'inflation unique, de manière indépendante les unes des autres. Cette approche, qui fut celle de Jevons puis d'Edgeworth, a été violemment critiquée par John Maynard Keynes, qui l'estimait hautement irréaliste. Walsh et Fisher, qui furent les deux plus grands théoriciens des indices de prix, ont tous les deux défendu des formules d'indices pondérées par les parts de dépense.

D'un point de vue méthodologique, l'indice de Laspeyres n'est pas parfait. L'emploi de pondérations basées sur la consommation en période de référence a comme conséquence que, plus on s'éloigne de ladite période de référence, plus les pondérations s'éloignent des « vraies » parts de dépense. On dit que l'indice de Laspeyres ignore les « effets de substitution ». En effet, si un produit devient fortement inflationniste, il est probable que les consommateurs chercheront à reporter tout ou partie de leur consommation sur un produit rendant un service comparable, mais moins inflationniste.
Dans la pratique donc et sur la plupart des produits de consommation courante notamment, un produit fortement inflationniste sera de moins en moins consommé, ce qui devrait se traduire dans l'indice agrégé par une moindre pondération de cette source d'inflation. Ce n'est pas le cas avec l'indice de Laspeyres. Toutefois, des études quantitatives ont pu démontrer que, dans le cas de l'IPC français, le biais de substitution était de faible ampleur. En outre, il n'est que rarement possible d'observer en temps réel l'évolution des quantités consommées, donc de connaître les parts de dépense de la période courante. Les « meilleures » formules d'indice, parmi lesquelles celles de Fisher ou de Walsh, qui utilisent des pondérations symétriques entre les périodes 0 et t, ne sont pas applicables dans le cadre d'une statistique comme l'IPC, qu'il est important de publier le plus rapidement possible afin qu'il puisse remplir son rôle d'indicateur conjoncturel.

Étant un indice de panier-type, l'indice de Laspeyres est également très facile à mettre en œuvre car il « suffit » d'observer les prix de produits prédéterminés, à intervalle de temps régulier, dans un même endroit. Des problèmes se posent toutefois lorsque les produits que l'on cherche à observer sont par nature volatiles, soit de manière saisonnière (c'est le cas de certains fruits ou légumes, des vêtements, des articles de rentrée scolaire, etc.) soit de manière définitive (c'est le cas des secteurs à fort taux d'innovation technique, par exemple). Des méthodes complémentaires doivent alors être mises en place pour s'adapter à ce profil particulier de produits.

Trois questions à Laurent Ferrara, Chef du Service de Macroéconomie Internationale à la Banque de France et Professeur associé à l'Université Paris Ouest

1) Dans le cadre de vos études et recherches à la Banque de France, est-il important d'identifier la tendance d'une série macroéconomique ou financière ?

Le service de Macroéconomie Internationale de la Banque de France s'occupe du suivi des pays industrialisés n'appartenant pas à la zone euro et de sujets plus transversaux comme les taux de change, les matières premières ou les déséquilibres mondiaux. Lorsque nous faisons le suivi conjoncturel de l'économie des pays ou des marchés financiers pour le Gouverneur de la Banque de France, il est important de dégager des messages clairs et de ne pas focaliser uniquement sur les derniers chiffres qui peuvent refléter des événements exceptionnels, voire inexpliqués. Dans un langage statistique et économétrique, cela signifie que nous cherchons à identifier le signal (c'est-à-dire la tendance et/ou le cycle) par rapport au bruit (ou composante résiduelle). L'extraction du signal est essentielle pour l'analyse économique car les relations macroéconomiques théoriques entre les variables sur lesquelles nous nous appuyons concernent en général le moyen terme, voire le long terme. Ainsi, filtrer le bruit de très court terme et récupérer les tendances de moyen et long termes constituent une grande partie du travail de conjoncturiste. Dans cette optique, les méthodes statistiques et économétriques sont d'une grande utilité, notamment les méthodes de lissage de type moyenne mobile ou les techniques de filtrage qui permettent de décomposer les variables macroéconomiques entre une tendance de long terme et une composante cyclique.

2) Quelles sont, parmi les séries que vous étudiez en macroéconomie ou en finance, celles qui présentent une saisonnalité marquée ? Pourquoi est-il nécessaire de les désaisonnaliser ?

La saisonnalité est souvent présente dans les séries macroéconomiques et vient « polluer » le diagnostic conjoncturel. En particulier, les séries de ventes ou de production sont fortement affectées par ce type de fait stylisé. Il semble en effet naturel que, par exemple, les ventes de voitures — une des variables les plus suivies pour l'analyse conjoncturelle — soient nettement plus faibles pendant le mois d'août que durant les autres mois de l'année. Inversement, la production de certains matériaux de construction, tels le béton, est très sensible au climat et a tendance à chuter pendant les mois d'hiver. Ces mouvements qui interviennent de manière périodique au cours d'une année doivent impérativement être pris en compte dans l'analyse. En effet, on ne peut pas comparer directement les ventes de voitures du mois de septembre avec celles du mois d'août. Une manière facile de tenir compte de la saisonnalité est de raisonner en termes de glissement annuel, c'est-à-dire de comparer la valeur d'un mois d'une année donnée à celle de ce même mois mais de l'année précédente. Cela est pratique, mais introduit un certain décalage temporel dans l'analyse. La littérature relative à l'économétrie des séries temporelles regorge de différentes méthodes pour décomposer efficacement ce type de séries entre une composante saisonnière et des composantes tendancielles et cycliques. La série filtrée de sa composante saisonnière sera appelée corrigée des variations saisonnières (CVS). Ce sont ces séries que nous privilégions pour l'analyse économique et nous faisons confiance aux producteurs de données (INSEE, Eurostat, Bureau of Economic Analysis, etc.) pour la qualité des données CVS qu'ils fournissent.

3) Quelles sont les séries sur lesquelles vous travaillez qui se prêtent le mieux aux méthodes de séries temporelles, c'est-à-dire aux méthodes consistant à expliquer une série par ses propres valeurs passées ?

L'analyse économique fait appel à un grand nombre de séries que l'on peut classer en trois grands groupes : données financières, qui reflètent l'activité sur les marchés financiers ; données issues d'enquêtes d'opinion (auprès des ménages, des industriels ou d'experts économiques) ; et données dites réelles (ventes, production, PIB,etc.). Parmi toutes ces séries, celles qui se prêtent le mieux aux méthodes de séries temporelles sont celles qui présentent la plus forte persistance (ou inertie). Typiquement, le taux de croissance du PIB ou de la production industrielle et les enquêtes d'opinion sont des séries caractérisées par une certaine persistance, rendant favorable leur modélisation *via* les méthodes de séries temporelles. Le taux de chômage et l'inflation sont deux autres séries qui possèdent une forte corrélation entre deux observations, permettant également le recours aux méthodes de séries temporelles pour leur analyse.Par exemple, le marché de l'emploi est connu pour présenter une forte hystérèse qui fait qu'à la fin d'une récession, le nombre de chômeurs continue d'augmenter alors que l'activité économique a repris, ce qui engendre ainsi de la persistance dans la série du taux de chômage. De l'autre côté du spectre se trouvent les variables relatives aux marchés financiers, telles que les taux de change, le prix des actions ou le prix du pétrole. Ces variables sont très volatiles et présentent très peu de persistance, les rendant difficiles à modéliser et à prévoir. En revanche, il existe une forme de persistance sur leur volatilité qu'il est souvent utile de prendre en compte. Au-delà de cette propriété de persistance, les séries macroéconomiques et financières présentent fréquemment des valeurs aberrantes, des ruptures, des non-linéarités, etc., rendant nécessaire la mise en place de méthodes statistiques et économétriques avancées pour aider à l'analyse économique.